Responses to Public Comments

Phase I: Review of Existing Methodologies

Phase II: Methodology Development and Derivation of Chlorpyrifos Criteria

of the

Methodology for Derivation of Pesticide Water Quality Criteria for the Protection of Aquatic Life in the Sacramento and San Joaquin River Basins

Amanda Palumbo,
Patti L. TenBrook, Ph.D.
and
Ronald S. Tjeerdema, Ph.D.

Department of Environmental Toxicology
University of California, Davis

With contributions from

Paul Hann and Danny McClure, P.E.
California Regional Water Quality Control Board, Central Valley Region

September 2009

THIS PAGE INTENTIONALLY BLANK

Table of Contents

Terms, Abbreviations, Acronyms, and Initialisms Used in this Report

Term	Definition
§	Section
ACR	Acute to Chronic Ratio- used to estimate concentration that will protect against chronic toxicity
AF	Assessment Factors- used to estimate concentration that will protect against acute toxicity when data set is too small to use SSD
CDFG	California Department of Fish and Game
CEQA	California Environmental Quality Act
CFR	Code of Federal Regulations
CRC	California Rice Commission
CVCWA	Central Valley Clean Water Association
DPR	California Department of Pesticide Regulation
EC50	The chemical concentration that has an effect on 50% of the test population.
IC25	The chemical concentration that inhibits a response in 25% of the test population.
ILP	Irrigated Lands Program
IRED	Interim Reregistraion Eligibility Decision
Koc	Organic Carbon Partition Coefficient
LC50	The chemical concentration that is lethal to 50 % of the test population.
LOE	Line of Evidence
LOEC	Lowest Observed Effect Level- lowest concentration tested that has some effect on the test population
MATC	Maximum Allowable Toxicant Concentration -geometric mean of LOEC and NOEC
MLOE	Multiple Lines of Evidence
NOEC	No Observed Effect Level- highest concentration tested that has no effect on the test population
NPDES	National Pollutant Discharge Elimination System
OPP	Office of Pesticide Programs, part of U.S. Environmental Protection Agency
Porter-Cologne	The Porter-Cologne Water Quality Control Act is the state of California's water quality control law.

PUR	Department of Pesticide Regulation Pesticide Use Report
QSAR	Quantitative Structure Activity Relationships
Regional Water Board	California Regional Water Quality Control Board, Central Valley Region
SSD	Species Sensitivity Distribution- Statistical probability distribution of toxicity data
State Water Board	State Water Resources Control Board
SWAMP	Surface Water Ambient Monitoring Program
TMDL	Total Maximum Daily Load
UC Davis	University of California, Davis
US EPA	U.S. Environmental Protection Agency
Water Quality Objective (WQO)	The limits of water quality constituents or characteristics that are established for the reasonable protection of beneficial uses of water or the prevention of nuisance within a specific area.
WDRs	Waste Discharge Requirements

1.0 Introduction

This document presents the responses to public comments received on two technical reports prepared by the University of California at Davis, Environmental Toxicology Department, under contract (#05-100-150-0) to the Regional Water Quality Control Board, Central Valley Region (Regional Board). These two reports represent the end products of two phases of a three-phase project to evaluate, develop and apply a method to derive pesticide water quality criteria for the protection of aquatic life.

The first phase of the project was to review and evaluate existing water quality criteria derivation methodologies to determine if there was an existing available method that met the Regional Board's stated project goals. If the review in Phase I had revealed a single methodology that contained all the features required by the contract scope of work, then that methodology could simply have been recommended as applicable for use by the CVRWQCB. However, the review indicated that there is no single method that meets all of the Regional Boards requirements. Therefore, the second phase of the project was to develop a new method that could meet the project requirements. The Phase II report details this new methodology and its application to chlorpyrifos. The new methodology is largely comprised of a medley of best practices from these other various methods, optimized to meet Regional Board Requirements.

Both the Phase I and Phase II reports were individually submitted to peer review, conducted by experts from academia and sister agencies, including the Department of Pesticide Regulation. The reports were revised to address peer reviewer recommendations and were then released for public review and comment between February and May 2007.

These technical reports may be considered by the Regional Board during the development of the Central Valley Pesticide Basin Plan Amendment or other Board actions. However, the reports do not represent Board Policy and are not regulations. The reports are intended to generate numeric water quality criteria for the protection of aquatic life. However, these should not be construed as water quality objectives. Criteria and guidelines do not have the force and effect of regulation, nor are they themselves water quality objectives.

While this is intended to be a technical support document, numerous comments extended beyond purely technical consideration and into policy consideration related to how the method would be used by the Regional Board. Beyond

defining the scope of the contract, the Regional Board has endeavored to separate policy issues from science issues in this project. Staff has taken several tangible steps, as described in response to comment 11-7, to prevent the appearance of policy directing science. To continue this separation, UCD and the Regional Board have generally supplied separate responses to several comments. Responses prepared by UC Davis are preceded by **UCD** while responses prepared by Regional Board Staff are preceded by **RB.**

2.0 Response to Comment to Public Comments

2.1. Comment Letter 1 – Roberta Firoved, California Rice Commission (CRC)

COMMENT 1-1: The CRC is unclear of the ultimate objective in developing methodology independent of the federal programs because the pesticide registration process provides aquatic toxicity evaluation under the environmental fate and effects review. The proposed project is an attempt to correct a process with no specific deficiencies. Staff have expressed highly speculative shortcomings in the current U.S. EPA methodology (EPA 1985) used for the data evaluation to register pesticides. Staff's concerns appear to be based on a misunderstanding of the pesticide registration process as considering only pesticide food residue tolerance.

RESPONSE TO COMMENT 1-1: **RB** Unquestionably, the registration process is an invaluable source for information about the environmental fate and effects of pesticides. However, the Regional Board cannot rely solely upon the current registration process to establish water quality objectives. The registration process does not generate a criteria value, which is required to meet the mandate of the Clean Water Act and Porter Cologne. Instead, the registration process determines whether the expected environmental exposures exceed a risk quotient that is acceptable to the Office of Pesticide Program[1]. If the expected exposure exceeds the "acceptable" risk quotient, mitigation or regulatory action may be taken, but the effects of the mitigation are not quantified to ensure that they are sufficient to prevent adverse effects to aquatic life. The recent Chlorpyrifos Interim Reregistraion Eligibility Decision (IRED) specifically acknowledges that even with mitigation, "potential risk to invertebrates, particularly estuarine invertebrates may still be of concern. Risk quotients represent a screening level assessment and are inadequate to predict whether the levels of chlorpyrifos entering estuarine areas are sufficient to affect invertebrate populations or populations of the larger species that depend on them as a food source." (U.S. EPA, 2002a). As a result, the Regional Board must look to another means to meet its requirement to establish water quality objectives to ensure the reasonable protection of beneficial uses.

[1] It is beyond the scope of this project to evaluate whether the risk quotient considered acceptable by the Office of Pesticide Programs would be acceptable by the requirements of the Clean Water Act and Porter Cologne.

Therefore, as stated in the project scope of work, the Regional Board has begun a project to identify or develop a method or methods for deriving numeric water quality criteria that are protective of aquatic life and could be used as the basis for pesticide water quality objectives in the Central Valley of California. The Regional Board and the California Department of Fish and Game have historically used the U.S. EPA Method 1985 methodology to derive water quality criteria. While this method is scientifically defensible, there are a number of reasons why the Regional Board is interested in investigating alternatives to the U.S. EPA 1985 method. These are as follows:

1. Available data on pesticides frequently does not include all of the taxa required by the U.S. EPA 1985 Method.
2. The U.S. EPA does not provide clear guidance when data for all eight required families are not available.
3. The methodology assumes an unbiased data set of toxicity results (i.e. toxicity tests were performed with no known or assumed tolerance of the species tested to the toxicant).
4. Stakeholders have suggested that more recent methodologies more accurately reflect current scientific thinking. Staff recognizes that other methods are available and believe it is the right time to evaluate them in a formal fashion.

The problem of lack of data is particularly difficult. As discussed in the Phase I report, the legal requirements for registering a pesticide may include as few as 3 taxa. In practice, data for more taxa are often generated; however, it is uncommon to generate sufficient data to use the EPA 1985 method. **Table 2-1** is a list of currently registered pesticides included on the 2006 TMDL list and pesticides identified through the draft Relative Risk Evaluation Technical report (Lu & Davis 2009) as having high relative risk. The table includes information about which pesticides have established water quality criteria or aquatic life benchmarks.

TABLE 2-1

ESTABLISHED CRITERIA FOR POTENTIAL PESTICIDES OF INTEREST

Pesticide	Reference [a]	Acute Criterion (µg/L)	Chronic Criterion (µg/L)	Benchmark Value Available [e]
Abamectin	None			
Azinphos Methyl	None [a]			✓
Bifenthrin	CDFG 2000	[c]	[c]	
Chlorothalonil	CDFG 1999	[c]	[c]	✓
Chlorpyrifos	Beaulaurier et al. 2005	0.015	0.025	✓
Cyfluthrin	None			
Cypermethrin	CDFG 2000	0.002 (Interim[b])	[c]	
Deltamethrin	None			
Diazinon	Beaulaurier et al. 2005	0.10	0.16	✓
Diuron	None			✓
Esfenvalerate	CDFG 2000	[c]	[c]	
Fipronil	None			
Hexazinone	None			
Lambda-cyhalothrin	None			
Malathion	CDFG 1998	0.43	[c]	✓
Mancozeb	None			
Maneb	CDFG 1999	[c]	[c]	
Methyl Parathion	US EPA 1986b	0.065	0.013	✓
(s)-Metolachlor	None			✓
Molinate	CDFG 1990	13 [d]	13 [d]	✓
Oxyfluorfen	None			✓
Paraquat Dichloride	None			
Pendimethalin	None			✓
Permethrin	CDFG 2000	0.059 (Interim[b])	[c]	✓
PHMB	None			
Propanil	None			✓
Propargite	None			✓
Pyraclostrobin	None			
Simazine	CA Primary MCL	4		✓
Trifluralin	None			✓
Tralomethrin	None			

[a] Where criteria have been set by more than one agency, the lowest applicable criteria has been listed. Criteria established prior to the EPA 1985 Method are excluded as they would require additional derivation work.

[b] Criterion is an interim criterion due to insufficient data

[c] The referenced report indicated that there was insufficient data to calculate criteria.

[d] The Criteria are absolute maximums

[e] U.S. EPA 2007b

31 pesticides are listed in **Table 2-1**, with less than half having some form of criteria established. Some of this is due to lack of study, but even where studied, there is a significant shortage of applicable criteria. Of the 12 pesticides on the list that have been studied by CDFG or the Regional Board, only 4[2] had sufficient data to use the 1985 method. Even Chlorpyrifos and Diazinon, which are generally recognized as having large data sets, have sufficient data to generate chronic criteria only through use of the acute to chronic ratio (ACR) procedure in the EPA 1985 method.

> **COMMENT 1-2:** Identifying deficiencies in aquatic toxicity warrants vetting at a federal level between the U.S. EPA/OPP and the Office of Water.

RESPONSE TO COMMENT 1-2: **RB** As discussed in the response above, there are differing requirements for the amount of data generated through the registration process versus the amount of data required to use the EPA criteria derivation process. Regional Board Staff agree that harmonizing these differing requirements at the federal level would be beneficial, and staff will continue to bring this issue to the attention of US EPA as opportunities become available. However, this will not remove the need for the Regional Board to meet its statutory mandate. Under Section 303 of the Clean Water Act, the state is specifically identified as having the mandate to establish water quality standards. Porter Cologne Section 13241 delegates this mandate to the Regional Board. To fulfill the Board's regulatory mandate, effective tools to establish water quality objectives are required. The methods employed by the U.S. EPA/OPP and the Office of Water are useful, but as described in response to comment 1-1, they cannot currently be used for all pesticides potentially of concern to the Regional Board. Therefore additional work to develop a suitable tool is required.

> **COMMENT 1-3**: Industry involvement is critical well beyond attendance at public workshops. The CRC has experience from the Rice Pesticides Program where industry input was instrumental in development of performance goals and water quality objectives to mitigate pesticide negative impacts while creating attainable numbers to those using the products.

[2] An early version of Table 3-1 was released at the May 2008 public meeting. The earlier table included more three more pesticides, including two that had criteria, but are neither 303(d) listed nor are rated as pesticides of high or moderate concern in the Relative Risk Evaluation.

RESPONSE TO COMMENT 1-3: RB As discussed in response to comment 1-1 above, one of the principal reasons we are evaluating new criteria methodologies is because this is exactly what was recommended by stakeholders during our previous diazinon and chlorpyrifos basin planning efforts. The Regional Board is interested in, and has provided opportunities for all interested stakeholders, including industry to comment on the proposed method. The workshops held as a part of this project have provided interested stakeholders with an opportunity to discuss their concerns not only with the Board, but also directly with the scientists involved in development work. In addition to public workshops, we are releasing each phase of method development for public review. Comments received on the methodology are being carefully reviewed and responded to by UC Davis (for technical issues) and the Regional Board (for policy issues). Any criteria derived using the new methodology will be reviewed in the Basin Plan Amendment Staff Report against a number of alternatives. The Staff Report will also be released for public review and comments. As we go forward with this phase of the project and eventually into preparation of the Basin Plan Amendment Staff Report, we will continue to provide interested stakeholders with numerous opportunities to affect the process and work products.

> **COMMENT 1-4**: The standards should address the specific toxicity concerns, but not at the lowest possible detection level - an over interpretation of current California water laws.

RESPONSE TO COMMENT 1-4: RB The purpose of the proposed project is to develop a tool to derive science based water quality criteria that are protective of aquatic life. The criteria developed using this methodology would be based on prevention of toxic effects, and not based on the lowest possible detection level. At this phase of development, the detection level is not considered in the criteria derivation.

It should also be noted that this methodology will only generate recommended water quality criteria, not Water Quality Objectives or “standards”. Within the context of this project, the term “criteria” refers to science based levels below which protection of aquatic life may be reasonably assumed. While criteria can be used in interpreting existing regulations, such as the narrative toxicity objective in the Basin Plan, they are not necessarily established in regulation.

In Contrast, Water Quality Objectives are firm limits established in regulation as part of the state’s Water Quality Standards in the Basin Plan. Objectives must be

adopted by the Regional and State Boards in a public process. Establishing objectives requires consideration of a number of factors beyond the scope of UC Davis's contract. Specifically, Porter Cologne Section 13241 requires consideration of the following factors:

- Beneficial Uses
- Environmental characteristics of the watershed
- Achievable water quality conditions
- Economic considerations
- Development of housing
- Development and use of recycled water

When adopting water quality objectives, the regional board will consider available criteria developed using this method as well as criteria developed using other methods. These will be compared against the Porter Cologne 13241 factors as well as other legal and regulatory requirements. This comparison will be documented in Basin Plan Amendment Staff Reports

Objectives can be set above or below derived criteria. Should a criterion derived using the proposed methodology be below the limit of detection using current methodology, the Regional Board could determine that any detectable amount of pesticide would prevent reasonable protection of beneficial uses. However, the Board could also that the objective should be set at some other level.

> **COMMENT 1-5**: The CVRWQCB should utilize the aquatic benchmark summary the U.S. EPA is developing as the basis in addressing pesticide concerns.

RESPONSE TO COMMENT 1-5: **RB** The Aquatic Life Benchmarks are useful screening values and may be considered as one of the potential alternatives in the Central Valley Pesticide Basin Plan Amendment staff report. However, the information provided with the benchmark values clearly states that the "benchmarks are only indicators, and such should not be confused with detailed risk and toxicity assessments that are conducted by the EPA in determining the safety of pesticides" (U.S. EPA 2007b).

Lacking other information, the benchmark values might be used to interpret narrative criteria, but would not be appropriate for adoption as numeric water quality objectives without additional investigation. Any investigation would be substantially equivalent to the work proposed in the methodology.

Only toxicity data from the registration process was used in deriving the benchmarks, instead of a broader literature search. Because other toxicity data were not considered, the benchmarks might not be fully protective of all aquatic life. A good example of this is the chlorpyrifos benchmarks. The benchmark for aquatic invertebrates is 0.04 ug/L. This is significantly higher than the current acute water quality objective of 0.025 ug/L derived using the EPA 1985 methodology. As compared to the toxicity data gathered for the current report, the benchmark value is higher than the lowest acute value of 0.035 ug/L for Daphnia ambigua and is equal to the lowest chronic value of 0.040 ng/L for Ceriodaphnia dubia in the current data set. Clearly, there is potential for harm if the concentration of chlorpyrifos were at the level of the benchmark.

Finally, as noted in **Table 2-1**, benchmark values are only established for a little over half of the pesticides that were identified through the draft Relative Risk Evaluation technical report (Lu & Davis 2009). For example, there are no criteria established for any of the pyrethroids except pyrethrin. Since the aquatic benchmarks have not been released for all of the relevant pesticides, there would still be a need for the methodology.

> **COMMENT 1-6**: Creating standards completely independent to California adds another layer of regulation and cost for doing business in the state. The additional costs create an economic disadvantage to California small businesses when the standards are not applicable on a federal level.

RESPONSE TO COMMENT 1-6: **RB** Under Section 303 of the Clean Water Act, establishment of water quality standards is specifically delegated to the state. Under Porter Cologne Section 13241, this authority has been specifically delegated to the Regional Water Boards. Where available, federally derived criteria will be considered as alternatives for adoption as a water quality objectives. However, as noted in **Table 2-1**, federal criteria are unavailable for many of the pesticides identified through the Relative Risk Evaluation technical report (Lu & Davis 2009). Economic costs will be evaluated as part of the Central Valley Water Pesticide Basin Plan Amendment Staff Report.

2.2. Comment Letter 2 – Dee Ann Staats, Croplife America

COMMENT 2-1: Croplife America (CLA) is concerned that this new methodology and its anticipated use has the potential to effectively insert this Regional Board into the establishment of pesticide use criteria and restrictions in a manner that effectively bypasses and potentially duplicates the existing registration, labeling and federal water quality regulatory structure for these products.

RESPONSE TO COMMENT 2-1: **RB** The registration process and our process are completely separate. Registration allows the pesticide to be used and the label explains how to apply the chemical. Our water quality objectives are about the amount that may be discharged into ambient waters. We are not prohibiting the use nor are we providing directions on its use. The authority to regulate the use of pesticides rests with the Department of Pesticide Regulations. Coordination with DPR is described and governed by a Management Agency Agreement (MAA). Regional Board Staff are coordinating with DPR Staff and DPR Staff are part of the team that peer reviewed the methodology prior to public release. Use of any derived criteria or management activities proposed to meet water quality objectives will be reviewed for consistency with the MAA as part of the Basin Plan Amendment Staff Report.

COMMENT 2-2: CLA is also concerned that the Regional Board is taking on the formidable task of developing what appears to be a new national criteria derivation process.

RESPONSE TO COMMENT 2-2: **RB** The Regional Board has neither the authority nor the intent to establish a national criteria derivation process. However, Regional Board Staff believe that the work being performed by UC Davis is valuable research, and may be of interest to others. As such, the scope of work includes publication in a peer reviewed journal or other equivalent publication. Once published, it will be available to any interested party.

COMMENT 2-3: The process has not identified deficiencies in the methods used by U.S. EPA and the California Department of Fish and Game.

RESPONSE TO COMMENT 2-3: **UCD/RB** See response to comments 1-1.

COMMENT 2-4: We believe the role of the Regional Board should instead focus on how available tools can best be applied or adjusted to take into account the site specific or regional ecosystem characteristics found in the Central Valley.

RESPONSE TO COMMENT 2-4: **UCD** Section 3-5 and 3-6 of method includes considerations on water quality affects, bioavailability, endangered species, mixtures, etc., all of which can be used to tailor criteria to Central Valley conditions, depending on the information available. Factors such as dissolved solids, temperature, presence of endangered species or other commonly used pesticides could be considered in the final criteria or on site-specific basis if appropriate.

RB The proposed project is essentially equivalent to what is suggested by the commenter. Specifically, the stated purpose of the Phase I report was to review and evaluate existing water quality criteria development methodologies ("available tools") to determine if there was an existing available method that could be used by the Regional Board. If the review in Phase I had revealed a single methodology that contained all the features deemed important to derivation of protective criteria, then that methodology could simply have been recommended for use by the CVRWQCB. However, the review indicated that there is no single method that meets all of the Regional Boards requirements. The new methodology is largely comprised of a medley of best practices from these other various methods, optimized to meet Regional Board Requirements.

COMMENT 2-5: The protection goal is not defined. The widely accepted concept that aquatic ecosystems can tolerate some stress and therefore protection of all species always and everywhere is not necessary (U.S. EPA 1985 method) is not discussed in the context of protection goals that will be met by the new proposed method.

RESPONSE TO COMMENT 2-5: **UCD** As required by the Project Scope of work, the aim of this method is to extrapolate from available toxicity data for a limited number of species, to a concentration that should not produce detrimental physiological effects in aquatic life. These criteria aim to protect all species in the freshwater habitats in the Sacramento-San Joaquin watersheds. This goal has been clarified in the methodology through the addition of a new Section 3-1.0 Goals and Definitions.

Section 2-1.1 has been revised to describe how the disappearance of a single species could lead to the unraveling of community structure due to complex interactions among species, suggesting that ecosystems might not be fully protected if water quality criteria are derived by a method that does not have the goal of protecting all species. Therefore, the new method has the goal of protection at the species level in order to fully protect natural ecosystems and meet the policy mandate. A detailed discussion of the project goals of other methodologies is provided in Section 5.0 of the Phase I report.

RB To meet the requirements of the Regional Boards legal mandate, the protection goal must be consistent with the requirements of Porter Cologne section 13241, which requires providing reasonable protection for beneficial uses. As discussed in the project scope of work, the protection goal must be protective of aquatic life beneficial uses as described in the Basin Plan, including but not limited to use of water that support preservation or enhancement of aquatic habitats, vegetation, fish or wildlife, including invertebrates. The protection goal must also be consistent with existing water quality objectives. The two most relevant objectives are the narrative objectives for pesticides and toxicity. The pesticide narrative objective requires that pesticide levels not be present in concentrations that adversely effect beneficial uses. The narrative objective for toxicity states that "all waters shall be maintained free of toxic substances in concentrations that produce detrimental physiological responses in …aquatic life."

> **COMMENT 2-6**: The need for a new method is not explained. Existing methods are capable of dealing with both robust and sparse toxicity data sets.

RESPONSE TO COMMENT 2-6: **UCD/RB** See responses to comments 1-1

> **COMMENT 2-7**: Specific procedures in the new proposed method exclude data that may have been used previously in existing methods, resulting in greater uncertainty in the final acute and chronic criteria for pesticides, and failing to consider the best available data.

RESPONSE TO COMMENT 2-7: **UCD** The new methodology attempts to use the best available data, and applies defined data quality guidelines. This makes the data evaluation process far more objective than in the currently used EPA 1985 method. Using less data increases the uncertainly in a criterion, but using low quality data has the same result. This method tries to find a balance between

using the most data and only best quality data and makes the process transparent and objective. A comparison of the data used by previous methods has been included in Chapter 4.0. See also response to comments 4-1 and 4-2.

> **COMMENT 2-8**: Making a rigid recommendation to take action in all cases when any level of exceedance occurs above a highly protective criterion value more than once in a three-year period is not scientifically justified.

RESPONSE TO COMMENT 2-8: **UCD** The scientific basis for the recommended frequency of exceedance is thoroughly discussed in Chapter 2-3.4 of the Phase II Report. This discussion concludes that 3 years between exposures should allow full recovery from effects of an excursion above either acute or chronic water quality criteria in the Sacramento and San Joaquin River basins.

2.3. Comment Letter 3 – Warren Tellefson, Central Valley Clean Water Agency

> **COMMENT 3-1**: CVCWA is concerned that the proposed methodology creates a process to rely on fewer data points to establish water quality criteria. The establishment of water quality criteria must be a scientifically robust process that includes a high quality and robust data set. The derivation of criteria with less data may result in lower water quality criteria that are overly conservative.

RESPONSE TO COMMENT 3-1: **RB** The Regional Board's mandate from Porter Cologne Section 13241 is to establish water quality objectives that are protective of aquatic life uses. Staff agrees that derivation of water quality criteria in support of this mandate would ideally benefit from a large dataset. However, the current pesticide registration requirements do not require sufficient data to use the EPA 1985 methodology and in practice, such datasets are not generated (see response to comment 1-1). As a result, the Regional Board must have a means to make the best use of available data. Therefore one of the goals of the method is to develop a method that is applicable to data sets of varying quantity.

The alternative is for the Regional Board to continue to refer to available data and information on a case by case basis to interpret the narrative pesticide and toxicity objectives. Under this alternative, the Basin Plan already provides for interpretation of narrative criteria using just the single lowest LC50. The proposed method provides an alternative to the current policy in that it allows

consideration of all available data by using a variable safety factor that is based on a statistical analysis of known datasets. In addition the method is written so that this number can change as additional information is obtained.

> **COMMENT 3-2**: The Regional Water Quality Control Board ("Regional Board") is developing a process that equates to the establishment of water quality objectives. Thus, the Regional Board is legally required to comply with sections 13241 and 13242 of the California Water Code for each criteria/objective that is derived from this methodology. At this time, CVCWA has yet to see how the Regional Board intends to comply with Water Code sections 13241 and 13242 when utilizing this methodology.

RESPONSE TO COMMENT 3-2: **RB** The proposed methodology is being developed to generate water quality criteria, not water quality objectives. In the context of this project, water quality criteria refers to levels that should be protective of aquatic life. They are science based numbers which are not intended to take into account the requirements of Porter Cologne 13241 and 13242. In contrast, water quality objectives are legal limits on the amount a given pollutant within a water body. At such time as the Central Valley Water Board determines to adopt Water Quality Objectives based on the criteria, those objectives would be subject to Porter Cologne 13241 and 13242.

2.4. Comment Letter 4 – Nick Poletika, Dow AgroSciences,

> **COMMENT 4-1**: The authors excluded all of the Good Laboratory Practice (GLP) studies submitted to U.S. EPA in their data analysis, claiming that the U.S. EPA's review practice is unreliable and that the reports for the studies are not available for their review. They use this reasoning to exclude critical studies such as the Daphnia magna reproduction study and the fathead minnow full life cycle study from use in setting the chronic criterion.
>
> The logic of the authors for excluding such data is faulty for the following reasons.
>
> - Registrant studies are held to a very high standard of data quality by the need to follow standard study guidelines[3] and GLP

[3] 40 CFR Part 158 Data requirements for registration, § 158.20(c) Availability of related guidelines.

requirements[4] independent of any subsequent review process. Therefore, the internal U.S. EPA review process builds upon the guideline study protocols and the GLP documentation and auditing process, thus insuring a high baseline of study quality. The U.S. EPA carefully reviews each study to judge acceptance or rejection based on meeting the protocol requirements and validation criteria. There are a total of 30 data points for aquatic species in the U.S. EPA database[5] on chlorpyrifos that are rated "Core" studies by U.S. EPA that have been ignored in this review.

- All of these data are available from U.S. EPA or Cal EPA Department of Pesticide Registration.

RESPONSE TO COMMENT 4-1: **UCD** Requesting studies from EPA and DPR are now the first suggestions in the table of data sources (Chapter 3, Table 3-1) as they are recognized as good information sources, but can take a long time to receive the information. On the other hand, original data found for several studies found in the OPP database (http://www.ipmcenters.org/index.cfm) did not pass the data screening process, indicating that reliability and relevance of data are not ensured merely by inclusion in the OPP database. Some of these studies are also not included in EPA own criteria reports. As discussed in response to comment 2-7 the new methodology attempts to use the best available data, and applies defined data quality guidelines.

The authors revisited the review of the Chlorpyrifos 2002 Interim Registration Eligibility Decision (IRED, US EPA 2002a) and found additional studies that had not been included in the draft report and these were requested and incorporated. Two fit the description of the studies mentioned by the commenter. One of the studies for fathead minnow (Mayes et al, 1993) rated highly and has been included, although, there has been no change in the calculated criteria. There are not enough data to perform a SSD with the chronic values. Therefore, the acute to chronic ratio was used. The study in question did not include an appropriate acute value (nor was an appropriate value found in another study), so it cannot be used in the acute to chronic ratio calculation. During the initial draft of the report, the fathead minnow results were excluded not because of any data quality issues, but because the original data were incorrectly believed to be unavailable. The author did not pursue the issue further because the toxicity values were redundant with fathead minnow data from an EPA lab already judged relevant

[4] 40 CFR Part 160.
[5] http://www.epa.gov/oppefed1/general/databasesdescription.htm#ecotoxicity

and reliable that was included in the acute to chronic ratio. The author recognized that obtaining this study would have no effect on the final criteria.

The other study requested was the only chronic study for daphnia magna in the OPP database (McCann, J. 1979). This study was not rated relevant and reliable. This is primarily because the study reported a backwards dose-response, in which survival and reproduction improved with higher chlorpyrifos of concentration. Only the one highest concentration tested had a lower response than the control. The low control response compared to other treatment is sign of a problem during the test. It is not clear that the NOEC and LOEC are from any statistical analysis and original data is not provided to do statistical analyses. The report does not describe the test type, but a citation for the study states it was a static test. The 21-day test is too long not to have some renewal of solution or some monitoring of water quality parameters, such as dissolved oxygen. No water quality parameters were reported including temperature. The lack of basic information in the 1 page report and anomalous responses make this test unsuitable to include in the criteria calculation. This study predates EPA's 1986 chlorpyrifos criteria report (U.S. EPA 1986a) and was not included in that report either.

Other acute studies for *Hyalella azteca*, *Daphnia magna Lepomis macrochirus Oncorhynhus mykiss* found on the OPP database were also included in the revised chlorpyrifos criteria report. A complete comparison of the data used by this and previous methods has been added to Chapter 4.0.

> **COMMENT 4-2**: The newly-developed final acute and chronic criteria for chlorpyrifos are lower than those established by the U.S. EPA and the California Department of Fish and Game (CDFG). These differences are explained by the authors as due to different data sets used for the final calculations. In view of our previous comment regarding the exclusion of GLP studies submitted and accepted by the U.S. EPA, DAS recommends a detailed comparison be developed showing the specific data sets used to calculate the U.S. EPA, CDFG, and the proposed new criteria. Without this additional information, it is impossible to distinguish between data selection and algorithms used to manipulate the data to understand their relative influence on determining the final numbers.

RESPONSE TO COMMENT 4-2: **UCD** A detailed comparison has been added to Chapter 4 as appendix C. Although the data sets are somewhat different, the rational for data selection are fairly comparable among three agencies compared: EPA, California Department of Fish and Game (CDFG, Siepmann & Finlayson

2000) and RB. For instance, in the methods of all agencies, studies were excluded if the control response, test procedures, test material, or dilution water was not adequately described.

Still, there were differences in the resulting data sets. These mostly come from:

1) Inclusion of newer data that often included more sensitive species. In particular for chlorpyrifos, some of the very lowest values included in the RB document were all published recently, including those for *Hyalella azteca* (Anderson & Lydy 2002), *Simulium vitatum IS-7* (Hyder et al. 2004), *Procloeon sp.* (Anderson et al. 2006). Overall, the RB data set includes values from 22 studies published after 1986, when the U.S. EPA criteria were derived (U.S. EPA 1986a), and values from 9 studies published in or after 2000, when the CDFG Sacramento and San Joaquin River criteria were derived.

2) The exclusion of values (often larger acute values) because the studies were not documented well. These studies lacked a description of the control and control response, some water quality parameters, the concentrations used and other information). The other agencies had knowledge of these studies beyond what was published (cited personal communication) or that they assumed they were conducted well because of the reputation of the laboratory or because the study relied on a citation to American Society for Testing and Materials (ASTM) methods without sufficient supporting documentation. The sources of most of this data are two large volumes of toxicity data; Mayer and Ellersieck (1986) and Johnson and Finley 1980. Mayer & Ellersieck 1986 contains most data in Johnson and Finley 1980. Mayer & Ellersieck 1986 assumed all tests met cited ASTM and US EPA methods. Johnson and Finley 1980 describes methods in detail, but not use of controls. To be fair and impartial in rating quality of all studies, one should refrain from making such assumptions and evaluate only information reported or obtainable. Additionally, in order to ensure reliability of data, the original reports should be reviewed.

Other notes from CDFG indicated that they tried to contact labs to obtain needed information where the published study was lacking. RB has contacted one lab to supplement methods information in one of the studies used. However, contacting someone from an older resource with so much data seems unlikely to produce the missing data. Studies in Mayer & Ellersieck 1986 were conducted between 1965 and 1984, so obtaining such information was not considered a viable option.

COMMENT 4-3: Finally, the authors did not interpret the evidence from mesocosm, microcosm, and ecosystem studies appropriately. Instead of considering whether the protection level of the proposed criteria agreed with the results from these studies, the authors merely stated that the proposed criteria will be protective. Since the authors do not specify in the methodology their protection goal, such a comparison has little meaning. In contrast, when the population and community level information is evaluated and then compared to other data in a multiple lines of evidence approach, it is clear that the levels needed to protect aquatic communities are much closer to the U.S. EPA acute criterion than the CDFG or proposed new criteria, and chronic effects with chlorpyrifos typically are not observed in aquatic ecosystems.[6]

RESPONSE TO COMMENT 4-3: **UCD** The goal of the method is to estimate criteria that are protective of all species in an ecosystem (see response to comment 2-5). The purpose of the method is to calculate protective values to prevent adverse physiological effects in aquatic organisms, not to prove impairment. Additional lines of evidence are not required by EPA or other agencies in setting water quality criteria and they are not necessary to derive protective estimates.

The method does call for comparison of criteria to field and mesocosm data. All such data that are of acceptable quality are checked over to see if there is any species that appears to not be protected by the current calculated criterion. If toxicity values obtained for appropriate endpoints (i.e., those related to survival, growth or reproduction) in high quality multispecies studies are lower than the derived criteria, then criteria may need to be adjusted downward. The recommended means of making a downward adjustment is to use either a lower 95% confidence limit estimate of the 5^{th} percentile (see discussion in section 2-3.1.3), or a median or 95% confidence limit estimate of the 1^{st} percentile. There is no recommendation to adjust criteria above these single species values this may lead to toxicity to the most sensitive species.

Additionally, single species values are doubled checked in the same manner. This is partly done just as an easy check for errors calculation. Other than errors, if there is good evidence that the criteria may not be protective of a species (in an acceptable quality test with measured concentrations) then the criterion should be lowered so that it is protective.

[6] Giesy, J. P., K. R. Solomon, J. R. Coates, K. R. Dixon, J. M. Giddings and E. E. Kenega. 1999. Chlorpyrifos: Ecological Risk Assessment in North American Aquatic Environments. Rev Environ Contam Toxicol 160: 1-129.

Mesocosm, microcosm, and field data is not used directly in the calculation because it is problematic due its paucity, variability and because they often lack clear dose response information. In field studies multiple chemicals are present and it may be impossible to determine where effects come from. In contrast, single-species toxicity tests can be successfully used in various extrapolation procedures to determine concentrations that are protective of ecosystems (see section 2-2.1.4).

Mesocosm, microcosm and fields data information is not enough to justify adjustment upward. Essentially, mesocosms and microcosms are vastly simplified ecosystems and should not be used as being synonymous with ecosystems, or population and community level responses. Although they incorporate parameters not included in standard aqueous toxicity tests, the individual components and exposure conditions have to be carefully reviewed to evaluate their realism. These data cannot automatically be assumed to equate to a more realistic toxicity value that should override single species laboratory tests. Moreover, because mesocosm and field studies are large scale and expensive tests, they often lack replication, in which case effects may be specific to the one mesocosm or field site tested. Additionally, it must be remembered that field sites and microcosm and microcosms may not contain sensitive species or be in a comparable ecosystem. These data will not be used to adjust criteria up above single species LC50 data based on field/mesocosm data as this may lead to toxicity to the most sensitive species. Section 6.4.3.3 of the Phase 1 report, which discusses mesocosm data in detail, has been incorporated into the Phase 2 report (section 2-2.1.4).

Further, the protection goal of the proposed method isn't at the functional level, but the species level. Some mesocosm, microcosm or field studies determine an ecosystem functional no-effect level based on the principle that if a species is lost but another is there to assume that function then the overall ecosystem function is preserved and thus is a no-effect concentration for the compound (see comment 14-44 for more details). This is not safe to assume, as a one species may be eliminated that causes ripple effect, changing ecosystem structure and ends in a loss of another species that is perceived as irreplaceable or is more economically or recreationally important.

To meet the stated goal of the criteria (protective of all species) it becomes necessary to adjust the criteria if a known toxicity value is lower than the criteria concentration derived by the proposed methodology. Ideally, there should be

enough data to eliminate the need for any kind of “post-hoc” adjustment. It is then up to the board whether or not to implement such criteria, as well as the decision-making on impairment.

RB Several commenters suggested that a MLOE framework is required to verify that additional lines of evidence are needed to determine whether specific local aquatic communities are actually impaired when any numeric criteria are exceeded. This comment seems to pre-suppose that there is a problem only when an entire community is affected as opposed to the individuals within the community. Such an interpretation would be inconsistent with the existing narrative toxicity objective, which requires prevention of detrimental physiological responses in aquatic organisms. See response to comment 2-5 for additional discussion on the protection goal.

Even if an entire community needs to be impacted in order to be impaired, the point of water quality criteria is not to get to the level of impairment. So, since we don’t want to see specific local aquatic communities be impaired when the water quality criteria are exceeded, an MLOE framework is not required.

2.5. Comment Letter 5 – William Thomas, Dow AgroSciences

COMMENT 5-1: On what basis does UC and the RWQCB believe the existing scientific reviews conducted by U.S. EPA, California DPR, other researchers and scientific authors are wholly inadequate and in need of rejection/abandonment?

RESPONSE TO COMMENT 5-1: **UCD/RB** Please see responses to comments 1-1.

COMMENT 5-2: The UC report seems to exclude considerable data in their review which appears to serve their desire for further restrictive regulatory application even though this departs from adopted scientific practices, skews data, increases uncertainty, decreases data robustness and departs from relying on the best scientifically available data.

RESPONSE TO COMMENT 5-2: **UCD** See response to comment 2-7, 4-1 and 4-2.

COMMENT 5-3: The UC report over relies on the premise of one exceedance in three years in that EPA only uses that threshold factor for industrial chemicals, not crop protection chemicals.

RESPONSE TO COMMENT 5-3: **UCD** This exceedance limit is derived from studies on the ecological effects of temporary contaminant exposures, and the time needed for ecosystems to recover. The review chapter 2-3.4 focuses on pesticides and this information was the basis for the 3 year time period in the UC report.

COMMENT 5-4: Specific to chlorpyrifos, the UC report totally and inappropriately ignored much of the extensive data base applicable to chlorpyrifos. Ignoring select data totally undermines the scientific credibility of the UC suggestion to further lower the chlorpyrifos threshold.

RESPONSE TO COMMENT 5-4: **UCD** See response to comment 2-7, 4-1 and 4-2.

COMMENT 5-5: The UC report seems to use a discounted reliability factor of 75% in evaluating data bases. Because chlorpyrifos has an extensive and robust data set this factor would be inappropriate.

RESPONSE TO COMMENT 5-5: **UCD** Evaluating the quality of scientific studies requires best professional judgment, which can be subjective. To make the process more objective the proposed method uses a numerical rating system, which is favorable because the rating the quality of studies will be more consistent and transparent. The 75% percentile was based on a review and comparison of many studies and is a good choice as verified by the comparison of different criteria (added to Chapter 4, see also response to comment 4-2). The system in the proposed method is not overly stringent compared to others. The advantage of the 75% cut-off is that it is always 75% and this can be much more objective than narrative guidelines for selecting which studies are to be used for the criteria calculation.

COMMENT 5-6: The UC reviewers seem to categorically ignore data from mesocosm studies which are among the most relevant data.

RESPONSE TO COMMENT 5-6: **UCD** See response to comment 4-3.

COMMENT 5-7: The present State Board review of sediment toxicity is focused on 1) multiple lines of evidence (MLOE) and 2) considerations of bioavailability, however, neither of these important scientific considerations are embraced in this report.

RESPONSE TO COMMENT 5-7: **UCD** Bioavailability and its effect on criteria derivation and interpretation are discussed in detail in Section 2-4.1. See response to comment 4-3 for discussions of multiple lines of evidence.

COMMENT 5-8: The Regional Board has just concluded its assessment of diazinon and chlorpyrifos and incorporated those new action levels along with the additivity formula into its Basin Plan and in three separate TMDLs. How can the Board justify initiating a separate and inconsistent assessment review within days of settling this issue in the Basin Plans.

RESPONSE TO COMMENT 5-8: **RB** Diazinon and Chlorpyrifos were chosen for two reasons. First, diazinon and chlorpyrifos have been the subject of considerable study. They have fairly large data sets and criteria have been derived at both the state and federal levels. As such, they represent good test cases against which to evaluate the new method.

Second, while diazinon and chlorpyrifos water quality objectives have been established for some of the waterbodies in the Central Valley, numerous waterbodies do not yet have established numeric criteria, including a number that are listed on the Current 303(d) list as impaired for diazinon or chlorpyrifos (see Table 2-2 below). Some of these impaired water bodies will be addressed through the proposed Basin Plan amendment. The California Code of Regulations (Title 23 §3777) requires that as part of the basin planning process that the Regional Board considers “Reasonable alternatives to the proposed activity.” This would include consideration of water quality criteria derived using the new methodology. In fact, chlorpyrifos criteria established using the new method were considered in the most recent Sacramento and Feather Rivers Diazinon and Chlorpyrifos Basin Plan Amendment, but were rejected at the time on the basis that the method was under development and the chlorpyrifos criteria were considered only preliminary.

COMMENT 5-9: Why has the Board engaged in a program which focuses only on chemicals used by agriculture to the exclusion of all other possible chemical contaminants.

RESPONSE TO COMMENT 5-9: **RB** The Regional Board is not focusing on solely agricultural pesticides. The choice of which pesticides to consider will be based in part on the Relative Risk Evaluation (Lu & Davis, 2009) together with data on what is detected through environmental monitoring as identified in the irrigated

TABLE 2-2

WATERBODIES IDENTIFIED AS IMPAIRED FOR DIAZION AND CHLORPYRIFOS AS INCLUDED ON THE 2006 303(D) LIST

Water Body Name	**Impaired for** [(a)]
Bear River, Lower (below Camp Far West Reservoir)	Diazinon
Butte Slough	Diazinon
Colusa Basin Drain	Diazinon
Del Puerto Creek	Chlorpyrifos & Diazinon
Harding Drain (Turlock Irrigation District Lateral #5)	Chlorpyrifos
Ingram Creek (from confluence with San Joaquin River to confluence with Hospital Creek)	Chlorpyrifos & Diazinon
Jack Slough	Diazinon
Main Drainage Canal	Diazinon
Merced River, Lower (McSwain Reservoir to San Joaquin River)	Chlorpyrifos & Diazinon
Natomas East Main Drainage Canal (aka Steelhead Creek, downstream of confluence with Arcade Creek)	Diazinon
Newman Wasteway	Chlorpyrifos & Diazinon
Orestimba Creek (above Kilburn Road)	Chlorpyrifos & Diazinon
Orestimba Creek (below Kilburn Road)	Chlorpyrifos & Diazinon
Salt Slough (upstream from confluence with San Joaquin River)	Chlorpyrifos & Diazinon
Stanislaus River, Lower	Diazinon
Tuolumne River, Lower (Don Pedro Reservoir to San Joaquin River)	Diazinon
Wadsworth Canal	Diazinon

[(a)] Listings of diazinon and chlorpyrifos impairments attributed to urban sources have been excluded from this table due to the cancellation of most urban uses of diazinon and chlorpyrifos in 2002.

lands reports and data submitted as part of the 2008 update to the 303(d) list. A preliminary list of pesticides that could be of interest is included in Table 2-1. As discussed in response to comment 1-1, the list in Table 2-1 was compiled based on the findings of the Relative Risk Evaluation (Lu & Davis 2009). Many of the pesticides in the Relative Risk Evaluation, such as bifenthrin and Cypermethrin were selected on the basis of the quantity of non-agricultural use. Others, such as Fipronil and Deltamethrin appear to be used nearly exclusively in non-agricultural settings.

COMMENT 5-10: The State Board is presently engaged in a focus on water sediment toxicity and in doing so is adopting a multiple line of evidence (MLOE) approach which this proposed program is apparently rejecting.

RESPONSE TO COMMENT 5-10: **UCD/RB** See response to comment 4-3 and 5-7.

COMMENT 5-11: There seems to be inadequate reliance on and coordination with California DPR and the Washington Office of Pesticide Programs of the U.S. EPA. There are jurisdictional, scientific, and consistency issues which should be addressed.

RESPONSE TO COMMENT 5-11: **RB** The Regional Board has been very careful to include the EPA and DPR in this project. Specifically, staff at DPR are participating as part of the peer review team for the project. U.S. EPA staff were also contacted to request that they join the peer review team, but declined citing potential conflict of interest issues. However, US EPA did submit comments on this method that will be considered in this response to comment document. Jurisdictional and policy consistency issue will be addressed in the Staff Report. See also response to comment 2-1.

2.6. Comment Letter 6 – William Warren-Hicks, EcoStat

COMMENT 6-1: The authors reviewed existing methodologies and finding them lacking, derived their own. However, Table 1.1 is incomplete and does not represent the current approaches for evaluating pesticide toxicity data. The table is focused on a selected few methods used by regulatory agencies rather than a compilation of the newest and most innovative methods. For example, the methods contained in the following examples do not have the statistical flaws contained in the subject document. There are many interesting and innovative methods for establishing criteria that are not referenced by the authors. Some recommended approaches are included in the following documents:

a. The U.S. EPA co-sponsored a Pellston conference entitled Application of Uncertainty Analysis to the Ecological Risk of Pesticides that produced many papers and approaches for establishing criteria. (In preparation for publication by SETAC Press, Pensacola, FL)

b. Species Sensitivity Distributions in Ecotoxicology. 2002. Leo Posthuma, Glenn Suter, Theo Trass. eds. Lewis Publishers, New York.

c. Warren-Hicks, W. J., D. Moore. eds. 1998. Uncertainty Analysis in Ecological Risk Assessment: Pellston '95. SETAC Press, Pensacola, Florida.

d. J. B. Parkhurst, Warren-Hicks, W. J., R. Cardwell, J. Volosin, T. Etchison, J. Butcher, S. Covington. 1995. Risk Managing Methods: Aquatic and Ecological Risk Assessment Aids Decision-Making. Water Environment & Technology, November 1995.

RESPONSE TO COMMENT 6-1: **UCD** This is one table in a large report that does not represent the entire contents of the report. It is an overview of major water quality criteria derivation methodologies currently in use, and shows major methods that address the whole process of water quality criteria derivation. Again, these are not the only approaches considered throughout Phase 1. Other ways of addressing the individual components to water quality criteria methodology are described later in the text.

After examining many sources, it was found that existing regulatory methods had elements that offered acceptable approaches. This is not all that surprising since those methods have been thoroughly reviewed by the corresponding agency. Many of the non-regulatory approaches would not work for the project purpose (e.g., too much data needed, models not quite accepted and validated yet), so if there was no better procedure found, the widely accepted, widely used approaches were chosen. The proposed methodology includes several elements that are not part of any regulatory method (mixtures, use of ACE and ICE programs, the rating system etc.), while the nuts and bolts (the SSD approach, the AF approach) are derived from existing regulatory methods.

Of the suggested references

- A is in preparation; not available.
- B is cited and discussed often in the Phase I report.
- C and D were reviewed and elements of these sources have been incorporated in revisions to the method (as described in response to comment 6-18).

COMMENT 6-2: The authors confuse concepts in statistics with concepts in laboratory testing. Methods used to analyze toxicity data should not be confused with quality assurance (QA) criteria that evaluate acceptable test results. Once laboratory data are deemed acceptable based on indicators of good laboratory practice, the data should be considered appropriate for statistical analysis. The results of a statistical analysis should not be used to judge the data invalid. For example, a high minimum significant

difference (MSD) does not necessarily reflect a “bad” test, but may reflect the fact that the toxicant is relatively weak resulting in a large between-concentration variance. A properly constructed MSD reflects the data; it does not invalidate the data as indicated by the authors.

There are a very few cases in which the output from a statistical analysis can be used as a QA tool for toxicological data. Statistical methods can be used as a tool to evaluate the presence or absence of the underlying conceptual model. For example, toxicological theory supports the concept that toxicity should increase with concentration. Statistical tests of monotonicity may be appropriate because they reflect the underlying model. [I note that monotonic dose-response curve issues are not considered here.] If the objective of the toxicity test is to identify say, an IC25, and the data are not sufficient for this purpose, then the test information may not be useful. Even in this case, however, providing the data meet basic laboratory QA criteria, the data themselves should not be labeled as invalid.

A large within- or between-concentration variance does not invalidate the data, nor does it invalidate the subsequent analysis of the data. QA decisions should not be solely based on the size of the variance estimator. It is important in this report that issues of quality assurance and statistics are not melded as they appear to be currently.

Response to Comment 6-2: **UCD** While the conditions used to perform tests and the statistical methods used afterwards are different things (quality assurance vs. quality control), there is no reason to keep them separate for this purpose. The reliability of test results can be judged based on both categories. Certainly, if a result is not statistically significant, that is a pretty good indicator that the results should probably not be used.

In the method's reliability rating procedure (Chapter 3, rating scheme, Table 3.8), a high MSD does not solely invalidate the data, nor is it a major factor preventing a study from being used in the criteria derivation. The MSD has a very small amount of weight the rating system. A low MSD is general a sign of reliable results. In that, it is given a very small amount of weight in the reliability rating system.1.5% of the reliability score depends on this parameter. The phrasing of the paragraph in Chapter 2-2.1. 2 has been revised to reflect that.

COMMENT 6-3: Note that model-based endpoints and hypothesis-based endpoints have many of the same interpretation issues. The authors state that: “Although regression methods are preferred, there is little agreement among scientists as to what level of statistical effect may be considered a

no biological effect ..." I note that this is the same issue surrounding the NOEC-LOEC-MATC issue that the authors discuss on the next page. The bottom line is that establishment of a statistical endpoint for setting criteria is arbitrary, whether the method is model-based or hypothesis based. I agree that model-based endpoints are preferred. It would be helpful if the authors could provide literature citations on this issue at this point in the document.

RESPONSE TO COMMENT 6-3: **UCD** Literature and citations on the use of model based endpoints were discussed in the phase 1 report section 6.4.3.2 Hypothesis tests vs. regression analysis. Some of this discussion has been added to the Chapter 2 report (2-2.1. 2).

COMMENT 6-4: It is also recommended that the authors point out in this section that model-based endpoints and hypothesis-based endpoints cannot be combined in the same analysis. These endpoints reflect a completely different underlying conceptual model. See more comments below on this issue.

RESPONSE TO COMMENT 6-4: **UCD** Section 2-2.1. 2 discuses the difficulty in equating the two endpoints. This is why hypothesis data is primarily used for the chronic data. However, it is reasonable to combine data from regression analysis with hypothesis generated data if species-specific studies are available to show what level of X (as in ECX or LCX) represents a biological no-effect level. Section 2-2.1. 2 has been clarified on this point.

COMMENT 6-5: The concept of a reliability score is inappropriate. See comments below. The rating techniques described in the ecotoxicity data evaluation section are inappropriate, have no viable interpretation, and do not represent acceptable quality assurance practice and procedure. Toxicity information, from multiple sources, should be judged based on the adherence to standard practice or guidance concerning the conduction of the test. Once the data pass these basic laboratory-based QA criteria, the data are then sufficient for statistical analysis. The weighting scheme described in this section is arbitrary, and the interpretation on p. 2-14 (based on percentiles) has no scientific basis. For example, why are relevance scores greater than the 90^{th} percentile rated "relevant"? Why not use the 85th or 95th percentile? This approach, based on questionable data management and poor statistical practice, is scientifically weak.

RESPONSE TO COMMENT 6-5: **UCD** The rating system does judge the adherence to standard practice or guidance concerning the conduction of the test. See response to comment 5-5 for more on reliability.

For relevance, the score of 90 or relevance was designed to inhibit the use of a study for criteria calculation that did not meet 6 very important and therefore heavily weighted requirements (endpoints linked to survival growth or reproduction, conducted in freshwater, chemical of at least 80% purity, family in North America, report a numerical toxicity values or one is calculable, report a control treatment/response). If a study does not meet one of these requirements the score would be reduced by 15 points. By this system only studies that have all 6 of these requirements will be used in the SSD and calculation of criteria. The lack of one or two of these important parameters would make the study available only for supplemental information (not used in the SSD calculation), and a lack of 3 more of these parameters excludes the study from the whole criteria derivation process.

> **COMMENT 6-6**: Note the following sentence, "The 75th percentile of scores is suggested for the reliability rating because, in the case of chlorpyrifos data set, higher percentiles were too restrictive, resulting in rejection of too much data ...." This sentence illustrates the issues I have with this section of the document. Good quality assurance criteria and procedures are not established based on the amount of data remaining for analysis. Data should not be discarded based on an arbitrarily established statistical endpoint. This approach to data reduction is inappropriate. Data should only be discarded if they are "wrong." An investigator may decided that data are not appropriate for the model under analysis, but the data should not be labeled "unreliable," since the data may be appropriate for other analyses.
>
> Again, the authors of this report are confusing good statistical practice with issues of basic data quality. This method results in a loss of information that may not be required from the perspective of identifying viable information. This section and associated procedures should be discarded.

RESPONSE TO COMMENT 6-6: **UCD** See response to comment 5-5 and 6-2.

> **COMMENT 6-7**: Sample size: Where did a sample size of 5 tests come from? A correct evaluation of sample size should consider the following: (1) the model under evaluation, (2) the variance of the data (or model terms), and (3) the pre-defined requirements for accuracy, precision, and

acceptable error. In this document, the sample sizes are established out of convenience. Approached scientifically, a “new” method would explore issues associated with precision, accuracy, and error requirements in a formal manner before blessing a sample size.

In any case, the objective of this document is a methodology for water quality criterion development. It is inadvisable to develop regulatory criteria from small data sets (n=5, for example). Therefore, based on good scientific judgment and practice, if data are lacking then the methodology should simply require additional data.

I strongly suggest that the authors conduct a formal sample size analysis before defining the number of toxicity tests required for criterion development. The authors should pay careful attention to issues associated with within- and between-laboratory variance associated with various toxicity test endpoints. There is a rather large literature on this subject, and again, a SETAC Pellston conference dealing with this issue.

RESPONSE TO COMMENT 6-7: **UCD** The analysis the commenter is calling for has been done by others and there are many references in the Phase I report. Much of this discussion has been now included in Ch 2-2.6. The sample size of 5 was selected by the developers of the distribution software, and other sources reviewed in Ch 2 led to the same number. A fit test has also been added to verify that the model fits each data set acceptably well. See also responses 6-11 and 6-10.

RB The need to develop regulatory criteria based on data sets of varying size is discussed in response to comment 1-1.

COMMENT 6-8: I understand the issue of over-weighting the relative toxicity associated with a specific species. In addition, I have no problem with the general guidance provided in the section on data reduction. I suggest, however, that the authors add narrative that addresses issues associated with combining the various toxicity endpoints (e.g., IC25, NOEC, LC50) within a single SSD. I strongly believe that a single SSD can only be comprised of data representing a single toxicity endpoint. This issue should be clarified and discussed in this section.

RESPONSE TO COMMENT 6-8: **UCD** It is clear in section 3-2.1.1.2 to only use LC50/EC50 data for the acute distribution. See response to comment 6-4.

COMMENT 6-9: The authors state that, “The aim of both SSD and AF methods is to extrapolate from available toxicity data for a limited number of species to toxicity values that will be protective of all species in an ecosystem.” Is this really the goal? If so, why not set the criteria to zero? Many regulatory agencies (e.g., the U.S. EPA) recognize that complete protection is not possible, or even useful.

RESPONSE TO COMMENT 6-9: **RB** As stated in the project scope of work and discussed in response to comment 2-5, the purpose of this project is to identify or develop a method or methods for deriving numeric water quality criteria that are protective of aquatic life and could be used as the basis for pesticide water quality objectives in the Central Valley of California. Regional Board Staff does not believe that meeting this goal requires setting pesticide water quality objectives necessarily to zero. Porter Cologne Section 13241 recognizes that it may be possible for the quality of water to be changed to some degree without unreasonably affecting beneficial uses, and provides a list of factors that must be evaluated in the process of setting beneficial uses.

COMMENT 6-10: Section 2-3.1.1 does not reflect good statistical practice and modeling. The stated reasons for selecting a distribution are inappropriate. Investigators should not choose a distribution because “... how many samples are required, on which distributions are easier to work with, or which ones better quell the criticism that SSDs are not valid ...” I note that this sentence represents a misunderstanding of good statistical practice. Also, I note that the U.S. EPA’s use of a triangular distribution, which has no interpretation within a toxicological paradigm, is also inappropriate. I encourage the authors to provide guidance for distribution selection that is consistent with good statistical practice.

Selection of an appropriate distribution should be based on the underlying conceptual model. For example, for acute data (life/death) binomial distributions of survival are appropriate. In this case, a generalized linear model using a log-logistic link function linking the concentration data to the probability of survival is consistent with the underlying conceptual model and the data collected to parameterize the model. Extensions of this model that lead to the mathematics underlying SSDs comprised of acute (binary) metrics can be found in the references provided at the top of this report. Distributions for continuous data or counts should be appropriately chosen. I refer the authors to a long series of papers written by A. J. Bailer and J. T. Oris for further examples of proper statistical approaches for developing models and distributions with toxicological data.

In no case should distributions be selected because they are convenient, easy to use, or match the available sample size (like the triangular). The

distribution must represent the toxicological and biological process of interest and be mathematically tractable within that process.

Selection of a distribution based on the conceptual model is a major reason why it becomes difficult to merge acute and chronic data into a single SSD. The endpoints are mathematically derived from different conceptual and mathematical models, making their cross-interpretation at best difficult, if not impossible. For example, it is difficult to compare an IC25 derived from an underlying binomial process with, say, a NOEC derived from an underlying log-normal process. Not only are the statistical methods used to generate these endpoints not comparable, the interpretation of toxicity inferred by each is incompatible.

As the field of environmental toxicology has matured, the acceptance of the link between distributions and biological interpretation has evolved. In the human health sciences this is equivalent to the interpretation of a gamma function for survival-time studies. Therefore, I see no need to involve another class of distributions (i.e., the Burr III family) without a thorough understanding of the link between the mathematics and biological and toxicological processes. I am not aware of published studies on how the Weibull and Pareto distributions are interpreted within the context of toxicological information. In fact, these distributions would be inappropriate for binary metrics.

I strongly suggest that this entire section be rewritten. The references at the top of my review have a great deal of information on how to correctly select and defend a distribution.

RESPONSE TO COMMENT 6-10: **UCD** None of the data used in the distribution are binary. Although the LC50's are based on life and death data, the LC50s are continuous data calculated with a linear model.

Different endpoints, such as IC25, LC50 and NOEC, should not be mixed and this is not done in the acute criteria distributions. Acute and chronic data are not merged into a single SSD.

Basing the choice of distribution on the underlying model nearly impossible in this case, see response to comment 6-11.

The above mentioned points were not the first factors used to choosing the distribution (“... how many samples are required, on which distributions are easier to work with, or which ones better quell the criticism that SSDs are not valid ...”). They were considered after fit and other considerations. Additionally, such practical considerations are not without value.

COMMENT 6-11: Choice of distribution should not be made based on statistical goodness of fit tests. The distribution must be interpretable within the process under evaluation. If the distribution does not fit the data, then the investigator may need to rethink his/her original hypothesis. But, the fit statistic should not be used to establish the underlying model (as is the case in this section). Please see my comments above. Also, I refer the authors to the above referenced documents on uncertainty analysis, which contain discussions of the misuse of curve fitting methods in decision-making. This section of the report should be rewritten to reflect the current literature in the statistical analysis of toxicological data within a risk paradigm.

RESPONSE TO COMMENT 6-11: **UCD** Finding a distribution that truly fits the exact underlying process is generally not feasible. This would require and very large amount of data and very good understanding of mechanism by which the compound in question affects all species, and more importantly why there are (or what are the) different sensitivities in different species. This information is not available for a great deal of species and chemicals.

With as much as is known about the underlying model this distribution is interpretable. Since there is a lot unknown about species variability a general model such as BurrIII that is applicable to many situations is appropriate.

The fit test was part of the decision, not the only deciding factor. This is a normal and commonly used approach when selecting a distribution for data for which the underlying model is not well characterized. There are no distributions specially designed for this data. Therefore, after making sure the distribution is applicable to these kinds of toxicity data (BurrIII family is a probability distribution that can incorporate concentrations (x values) from 0 to infinity, and it doesn't assume one distribution, but actually tests a few), it makes sense to choose on the fits the data well.

A step to verifying fit of the distribution to each individual data set with a fit test has been added in Section 3-3.2.4.

COMMENT 6-12: The authors should consider the statistical methods used in the Water Environmental Research Foundation (WERF) risk methods and software (cited above). The methods for distribution choice provided in the WERF documentation are consistent with good statistical

practice. Furthermore, the WERF method presents approaches for setting SSD-based criteria using uncertainty estimates. The method provides derivations of all statistical functions at fixed percentiles of the SSD, thus eliminating the need for safety factors.

RESPONSE TO COMMENT 6-12: **UCD** There was no software referenced and the WERF article referenced (in comment 6-1) was short and very general. The safety factor is used because the calculation uses LC50 data. The safety factor relates the LC50 to a no effect level based on data from many studies (see Comment 14-2). The distribution used will not to change the fact that the data used represent a concentration lethal to 50% of organisms. Any method that starts with LC50 data should have some way to estimate the no effect concentration. The BurrIII software also provides concentrations and uncertainties at any chosen percentile or vise versa.

COMMENT 6-13: The distribution of choice must reflect the biological and toxicological process under evaluation. If modalities are evident because of differing biological or toxicological processes, then I agree that data should be separated. They are separated because the underlying process is not consistent with the conceptual model reflected by the choice of distribution.

However, never eliminate data because they do not fit the model well. Valid data should be used, even if they do not represent the choice of distribution. Uncertainty in the data is informative and should not be eliminated. The author's statement that "... it is reasonable to exclude outliers ..." represents poor statistical practice.

RESPONSE TO COMMENT 6-13: **UCD** The method has been changed to remove the outlier test and box plot analysis. Critical examination of data is the only suggestion. If the distribution does not fit, one recommendation is to examine for bimodality and if apparent breaking into subsets in a justifiable manner (separating by taxa). If fit there is not enough data or a justifiable reason to break up data into subsets the assessment factor method is to be used.

COMMENT 6-14: Comparison of Methods: From a theoretical perspective, only the methods of Aldenberg & Jaworska (2000) overcome some of the many issues raised above. This method can be adapted for binary, continuous, and cardinal data. The other methods are lacking in either mathematical rigor, flexibility, or interpretation. The authors will find

that the Aldenberg & Jaworska approach is similar to the WERF methodology.

RESPONSE TO COMMENT 6-14: **UCD** The methods of Aldenberg & Jaworska do have an advantage in that information be input into the model to tailor it to the data set, but this also somewhat subjective and will be very influential with small data sets. Further, with small data sets there is not much information to base this prior information on. Therefore, a method that uses a general model that fits many data sets will work better, and this will avoid user bias of a model that can be tailored.

COMMENT 6-15: 12. p. 2-45, AF methodology: In this age of modern computing and internet communication, there should be no reason to use safety factors (a.k.a. assessment factors) - for any reason, under any conditions. The literature is replete with mathematical methods for calculating uncertainty in SSDs and concentration-response models. The use of safety factors is simply to assure policy makers that their resulting criteria are protective. However, as the narrative correctly notes (but for some reason then ignores), there is no safety in the use of safety factors. Their use is not a reasonable approach for predicting the future protectiveness of the criterion. Effectively, the use of safety factors negates the influence of science and mathematics in policy decisions.

Criteria should not be developed from small data sets. Therefore, the use of these factors is not justified based on first principles.

RESPONSE TO COMMENT 6-15: **UCD** The assessment factors are not arbitrary safety factors. They were empirically derived using mathematics and actual pesticide data sets. The factors were derived using a method that EPA used to derive factors for the Great Lakes methodology by Host et al. (1995). In the procedure, real data sets for pesticide were used, and randomly sampled to create subsets of the data that were used to simulate the case of having only 4 to 1 data points. Assessment factors were calculated from these numbers, as a way to estimate the same 5th percentile from the larger data sets when using only 4 values or less. .

More data are desirable, however a criteria based on limited data will do more to meet the protection goal than deriving no criteria. All criteria are estimates. These estimates are scientifically valid if the procedure and the assumptions involved are clear, justified, and use best means available. Science only indicates that there is a higher level of uncertainty when using few data. Science does not

stipulate that no criteria can be calculated with any amount of data. The proposed method includes the best means found that would meet the protection goal and mandate.

RB In the process of establishing earlier amendments, one of the common themes was that stakeholders generally prefer clear numeric objectives that specifically consider required Porter Cologne factors such as feasibility, costs and other values, and which are explicitly adopted by the Board. As discussed in response to comments 1-1 and 6-16, it is uncommon to generate sufficient pesticide data to use the current EPA 1985 method. As this situation is unlikely to change in the near future, an alternative method is required to establish criteria using a reduced dataset. The proposed scope of work therefore includes the requirement to be able to establish criteria for both large and small datasets. The method proposed by UC Davis for small datasets is an assessment factor method, which as discussed in the report is used by a variety of regulatory agencies, including the US EPA.

In addition, it should be noticed that existing Basin Plan policy already provides for establishment of criteria using just one toxic value. Under the current Basin Plan narrative policy, criteria are interpreted based on $1/10^{th}$ of the lowest LC50 value (equivalent to an assessment factor of 10) unless other criteria are already available, or until sufficient data is available to derive criteria using another method (typically the EPA 1985 methodology). In contrast, the proposed methodology allows consideration of all available data to derive criteria that are less likely to be over- or under-protective.

Finally, it should be clarified that the Technical report prepared by UC Davis generates criteria, not objectives. Criteria and guidelines do not have the force and effect of regulation, nor are they themselves water quality objectives. (see response to comment 1-4 for additional discussion on the difference between criteria and objectives).

> **COMMENT 6-16**: If the objective is to develop a water quality criterion with regulatory implications, then chronic data should be generated. Substituting the use of an ACR in place of generating chronic data is not appropriate for establishing standards with regulatory implications. The literature contains numerous studies (some referenced in the document) concerning the large range of ACR values for a single chemical/test species/endpoint combination. Therefore, ACRs should not be used, and chronic data should be generated when needed. Appropriate methods for

selecting the number of chronic tests is addressed in my comments above.

RESPONSE TO COMMENT 6-16: **UCD** The proposed methodology contains a procedure (similar to EPA's) that takes into account the possibility of a large range in species specific ACRs and the fact that sometimes much larger ACRs are calculated for species with the larger LC50s (insensitive species). ACRs from at least 3 families are required (see section 3-4.2.1 for full details). If the ACRs vary by more than a factor of 10 then all ACRs will not be used in the calculation. The ACRs that will be used are those from the species whose LC50(s) are closest to the acute criterion (which are usually smaller). This procedure is protective of the most sensitive species and minimizes unnecessary use of large ACRs.

RB The development of regulatory criteria using ACRs is established practice that has been used successfully by the EPA and other agencies for many years and is included in the EPA 1985 method. As discussed in response to comments 1-1, the current registration process does not typically generate sufficiently large datasets to perform SSD evaluations on chronic data. Even Chlorpyrifos and Diazinon, which are widely recognized as having large data sets, have sufficient data to generate chronic criteria only through use of the ACR procedure in the EPA 1985 method.

This situation is unlikely to change in the near future. The Regional Board does not have the authority to require pesticide manufacturers or registrants to generate and provide the necessary data. Neither does the Regional Board have the resources to conduct the research needed to fill the data gaps. Even if the Board did have the resources, for some priority pesticides, it could potentially take many years to collect the data that would be needed to fulfill EPA criteria development requirements. We do need to move forward with the development of TMDLs for problem pesticides. So, we need to be able to figure out how to work with the information that we have available to us now. This means figuring out how to work with relatively small data sets. The Basin Plan already includes a default system for working with a small data set. However the proposed method provides an alternative, based on more information, to the default system in the Basin Plan.

COMMENT 6-17: The statistical issues underlying an appropriate averaging period, and the number of water quality samples required during the averaging period, include the following: (1) temporal variability

in pesticide concentration, (2) occurrence of temporal correlation (i.e., autocorrelation or seasonal patterns), and the statistic of interest (mean, upper percentile, etc.). The U.S. EPA has regulatory guidance that was developed without the explicit analysis of these issues within the context of criterion setting. Approached scientifically a “new” method should address these issues when attempting to establish a time-period for water quality sampling. Without addressing these issues, the authors risk both false positive and false negative results within their regulatory framework. I suggest that the authors formally address the averaging time and associated sampling issues prior to endorsing an approach.

RESPONSE TO COMMENT 6-17: **UCD** This was discussed in Chapter 2. There are a lot of ideas out there for better ways to come up with averaging periods, but they're not very well developed yet. Time-to-event approaches are discussed in Phase I.

COMMENT 6-18: The authors should consider the WERF methods, and papers presented at the SETAC Pellston conference on Uncertainty Analysis of Pesticides, prior to selecting an approach to dealing with mixtures. These references present formal mathematical approaches for combining data across species within a single chemical, and across chemicals within a single species. The methods presented in this section lack a proposed approach for dealing with uncertainty. Also, these methods focus on a single effects endpoint rather than the entire concentration- curve. A “new” method should provide insights into more advanced methods for combining data.

RESPONSE TO COMMENT 6-18: **UCD** SETAC Pelleston reference is unavailable and the commenter did not supply the relevant references on the papers. The WERF reference is pretty general, but discusses the use of probability functions to estimate risk for each chemical. Similarly, the new method uses a probability function to estimate a concentration in which there is a low risk of harming aquatic life. It is likely that the biggest source of uncertainty is the unknown species to species variability estimated by the distribution. The proposed method is capable of providing the estimate with different levels of confidence to estimate the uncertainly in the extrapolation. We chose to start with the median estimate.

When working with possible mixture, the proposed method uses the resulting estimate with toxic equivalents or relative potency factors calculations, which are well accepted methods for working with mixtures. The software used by WERF methods, on the other hand, integrates the two probability functions. It is not

detailed how the two compounds are related to determine overall risk, but one would guess it is based on similar methods.

Different ideas on combining data are discussed in the Phase II report, but often we are faced with small data sets. It is also difficult to do a sound quantitative uncertainly analysis with limited data. Therefore reporting more deterministic protective concentrations is more appropriate.

However, after reading reference c from comment 6-1 (Warren-Hicks & Moore 1998), the authors have incorporated into the method more discussion of uncertainties and limitations. The method now includes an assumptions and limitations section that reviews some of the major uncertainties associated with data and procedures, assumptions and limitations of the method, and areas that would benefit form more research.

> **COMMENT 6-19**: The Burr III family is inappropriate for binary data. Furthermore, the Weibull and Pareto distributions have no interpretation within a toxicological context. This is comparable to the dangers associated with Monte Carlo software (a highly misused tool) where the investigator is free to select interesting distributions without any knowledge of the underlying mathematics or associated links to the biological process under evaluation.
>
> It is recommended that the authors provide sound mathematical and statistical arguments for their selection of the Burr III distributions. In particular, arguments linking the underlying biological process with the mathematics of the selected distribution should be provided. Currently, the document is lacking this defense. It is not enough to simply find an equation that generates a sigmoidal curve.

RESPONSE TO COMMENT 6-19: **UCD** See response to comment 6-11. The proposed method uses the BurrIII primarily for LC50 data, which are not binary data. These are estimates that can vary along a continuum. NOEC/LOEC data are not strictly binary either. The use of the Burr III family was chosen for the Australia/New Zealand method after much research into the distribution issue. It is difficult, especially with such limited data, to provide sound mathematical and statistical arguments linking the underlying biological process with the mathematics of a distribution. Again this is why one was chosen that is applicable to many situations.

COMMENT 6-20: The relevance scoring system should be discarded and replaced with sound quality assurance criteria and practice.

RESPONSE TO COMMENT 6-20: **UCD** See response to comment 6-5.

COMMENT 6-21: Excluding data simply because of fitting issues is poor statistical practice. Outlier tests (Sokal & Rolf) should not be used as the basis for discarding data that have passed rigorous laboratory derived quality assurance criteria.

RESPONSE TO COMMENT 6-21: **UCD** The outlier test was removed, see 6-13.

COMMENT 6-22: A formal analysis of the number of tests required for criterion setting should be developed and presented.

RESPONSE TO COMMENT 6-22: **UCD** See response to comment 6-7.

COMMENT 6-23: Assessment factors should be replaced with formal uncertainty analyses.

RESPONSE TO COMMENT 6-23: **UCD** The procedure used to derive the AF was a form of uncertainly analyses. See response to Comment 6-15.

COMMENT 6-24: It is recommended that the authors re-consider their approach to working with mixtures. The references provided above are useful in this regard.

RESPONSE TO COMMENT 6-24: **UCD** See response to comment 6-18.

COMMENT 6-25: A formal analysis of the statistical issues underlying the selection of an averaging period and sample size associated with water quality sampling should be developed and presented prior to publishing this document.

RESPONSE TO COMMENT 6-25: **UCD** See response to comment 6-17.

2.7. Comment Letter 7 – Stephen Clark, Pacific EcoRisk

COMMENT 7-1: Although Regional Board staff indicated that this approach was more protective than no numeric water quality criteria being established at all (i.e., RWQCB must meet their mandate to protect waters of the State), a more robust data set would surely provide a more scientifically defensible water quality criteria, as the LC50 generated from any one study may prove to be either overly sensitive or insensitive for a whole variety of reasons (e.g., acceptable Controls but sensitive batch of organisms, as demonstrated via a concurrent reference toxicant test).

RESPONSE TO COMMENT 7-1: **UCD/RB** The method is designed to use as much data as is available, provided it meets the data quality requirements outlined in the method. Please see response to comments 1-1 and 6-15

COMMENT 7-2: As the LC50 is obtained via a best fit line for a single dilution series exposure, this best fit line can readily be skewed to produce a hyper sensitive or hyper insensitive LC50; the LC50 could be significantly different if the study was repeated. There is no requirement in the literature rating system (Table 3.6) proposed by the authors to protect against this anomalous test result issue, or to have repeated measures (i.e., at a minimum, repeat the study to provide some measure of variability about the LC50) to assure that the I or 2 values that could be used in the AF approach actually are soundly produced values. Although protecting the waters of the State is in the best interest for all involved, clearly the water quality criteria should be scientifically defensible so as to justify any financial burden that may result for stakeholders.

RESPONSE TO COMMENT 7-2: **UCD** Repeated teats are desirable, but it is not the norm to have repeated tests and this point alone should not preclude the use of the data in question.

Repeated tests are not part of the scoring system, but they have more weight in the criteria. This is because the individual values are kept separate until the species mean value is calculated. For example if there were two minnow studies, one with one value and one with two values because they repeated the test, the geomean of the three values would be used. In that way the study with the repeated value does have more weight in the final species mean value.

Any studies used in the AF method would have to be rated highly by the current scoring system, and therefore would be considered soundly produced values.

COMMENT 7-3: The rating system should include a score for characterization of organism sensitivity (i.e., additional "points" for study with concurrent reference toxicant study)

RESPONSE TO COMMENT 7-3: **UCD** This would only make sense if sensitivity was compared to long term reference study results. This would add valuable data, but this information is rarely provided with study results. This is not a very feasible requirement for tests with many species, and so was not included.

COMMENT 7-4: The Rating System should include a score for repeated measures (i.e., study repeated to provide some measure of variability about the LC50).

RESPONSE TO COMMENT 7-4: **UCD** See response to comment 7-2.

COMMENT 7-5: Studies that use an impure chemical standard should not be included in the database used to generate a water quality criteria.

RESPONSE TO COMMENT 7-5: **UCD** This is already included in the scoring system. A study using an impure chemical standard would loose 15 points (out of 100) and studies with a relevance score of 70-90 are only to be used as supplemental information. So this instance (and those described in comments 7-6, 7-7, 7-8) would receive relevance scores of 85 and would therefore be unacceptable for criteria derivation by this method.

COMMENT 7-6: Studies that do not report toxicological endpoints of survival, growth, or reproduction (e.g., standard EPA endpoints) should not be used to generate water quality criteria.

RESPONSE TO COMMENT 7-6: **UCD** See response to comment 7-5.

COMMENT 7-7: Studies that do not describe the controls should not be used to generate water quality criteria.

RESPONSE TO COMMENT 7-7: **UCD** The rating system does not allow a study completely void of control to be used. The rating system is set up so that if

a study did not describe a control, but reported a control response (or the reverse) it would be acceptable, for the first rating anyway. The second scoring system also gives points based on use/ description of control. The loss of those points and points from other parameters could exclude the study. If rated highly otherwise, a study that had only a control description or only a control response would be used. Again, a study with no mention of control is not used.

COMMENT 7-8: Only data from freshwater organisms (versus salt water) should be used to produce water quality criteria for freshwater environments.

RESPONSE TO COMMENT 7-8: **UCD** See response to comment 7-5.

2.8. Comment Letter 8 – Claus Suverkropp, Larry Walker and Associates

COMMENT 8-1: The Risk Evaluation process is a very coarse screening procedure with some significant technical shortcomings. Specifically, the procedure relies largely on total pounds or pesticides applied and total acres as risk measures. These are not meaningful indicators of relative pesticide risk without considering application rates or effectiveness of the active ingredients. However, since it is used primarily as a screen to eliminate pesticides of relatively low concern, these shortcomings may not a major concern in the overall criteria development process. It is important to be clear that this is merely a screening process and that it is not really objective, since it relies on arbitrary judgments of relative risk. It should also be made clear that the Risk Rankings are not a regulatory assessment and only provide a focus for the criteria development process. In this context, the shortcomings of the ranking methods are less important. It should also be clarified that the final selection of pesticides for criteria development involves considerable subjectivity, with some pesticides being added based on "Best Professional Judgment" that otherwise would not have made the cut based solely on their more objective rankings.

As a follow-up to the Risk Assessment process, the Water Board should also verify whether "high use" and "high risk" pesticides (as defined by the Risk Assessment screening) have been detected in Central Valley surface waters at concentrations hypothesized to cause toxicity, and evaluate whether these concentrations actually caused toxicity.

RESPONSE TO COMMENT 8-1: **RB** This comment refers to other work being performed by Central Valley Water Board staff as a part of the larger Central Valley Pesticide Basin Plan Amendment. This comment will be addressed as part of the larger project.

COMMENT 8-2: The Aquatic Life Use Assessment is essentially based on evidence of presence/absence of any aquatic life. The main inputs for this evaluation were the stream names, bioassessment data from multiple sources, and critical salmonid habitat data from NOAA. The outcome of this assessment is that any natural stream with any evidence of any current or past aquatic life will be regulated based on Aquatic Life Beneficial Uses (e.g., COLD, WARM, migration, spawning). More than 700 named "Natural Streams" were identified in the Central Valley based primarily on naming conventions (e.g., river, creek, and slough vs. drain or canal). The evaluation is intended to exclude constructed agricultural drains, primarily because this was outside of the scope for the project. Because stream names were the only basis used to identify natural streams, Water Board staff should verify that the "sloughs" are natural streams, because "slough" has sometimes been applied to name waterbodies constructed for drainage.

RESPONSE TO COMMENT 8-2: **RB** This comment refers to other work being performed by Central Valley Water Board staff as a part of the larger Central Valley Pesticide Basin Plan Amendment. This comment will be addressed as part of the larger project.

COMMENT 8-3: The proposed criteria development process includes detailed guidelines for collection and review of the toxicity data. This appears generally to be reasonable and well-thought out. Overall, it provides for a structured and objective process to evaluate the data that are used in the criteria development. However, some elements of the evaluation will exclude data that were previously considered adequate for criteria development by U.S. EPA (as demonstrated by the chlorpyrifos example), and are useful in establishing the toxicity distribution. This will generally result in fewer data being used and therefore lower criteria (due to increased extrapolation at the sensitive end of the toxicity distribution).

RESPONSE TO COMMENT 8-3: **UCD** The proposed selection process for inclusion of data in criteria development is fairly comparable to the EPA method (see response to comment 4-2). Exclusion of certain studies used by other agencies may sometimes result in lower criteria, but it could also result in higher criteria. See also response to comment 2-7.

COMMENT 8-4: There are a number of technical problems with Criterion Derivation procedure that should be addressed before applying the criteria derivation process to pesticides. This is of additional concern because of the potential for the Water Board to use this same methodology to generate new criteria for non-pesticide parameters. There is nothing specific to pesticides in the criteria derivation procedure, and based on the proposed method, new criteria derived with this method can be expected to result in lower criteria much more often than not. When the potential for application of arbitrary safety factors (assessment factors or AFs) is combined with the relatively small amount of data available for most newer pesticides, it is a very likely that the proposed criteria development process will generate some very low and inappropriate pesticide criteria. Similarly, the provisions in the proposed method would result in lowered values for many other parameters which the Water Board may subsequently attempt to apply by replacing well-established criteria, or through interpretation of narrative criteria already in the Basin Plan.

RESPONSE TO COMMENT 8-4: **UCD** Criteria derived by this method will not necessarily be lower. A comparison of the proposed method to other methods has been included in Chapter 4.0 and is summarized in response to comment 4-2. The method is intended for pesticides. This has been clarified in the beginning of Chapter 3. The AF and default ACRs were derived from pesticide data only.

RB The goal of the method as stated in the Project Scope of Work and Chapter 1 of the methodology is "to develop a methodology for derivation of pesticide water quality criteria for the protection of aquatic life in the Sacramento River and San Joaquin River Basins." Uses of this methodology for non-pesticide parameters is not envisioned as there are aspects of the method that could not be applied without additional research. As a minimum, the default acute to chronic ratios and assessment factors are based solely on data for other pesticides and would have to be recalculated if the method were to be used for other pollutants.

Before a criterion derived using this method could be adopted as regulatory water quality objective, it would need to be evaluated in accordance with the requirements of Porter Cologne 13241. In the process of this evaluation and other relevant established values would also be reviewed as part of the alternatives analysis.

COMMENT 8-5: The proposed method explicitly provides procedures for derivation of criteria based on insufficient data. It accomplishes this by

requiring the use of arbitrary "Assessment Factors" (AF) to generate criteria for toxicity data sets with results for only one to five species. The problem of the lack of a valid scientific basis for specific AFs is trivial compared to the problem of generating criteria with insufficient data. Fewer than 4 toxicity values simply isn't a valid basis to derive a scientifically defensible criterion. EPA's minimum of 8 Genus Mean Acute Values (GMAV) or Species Mean Acute Values (SMAV) in this case, is a more reasonable amount of data - although still not adequate for accurate definition of the overall distribution, it provides a reasonable compromise between certainty and the cost of generating criteria. It was suggested by Water Board staff at an April 18, 2007 workshop that the low data requirement threshold (one value) serves to motivate regulated entities (e.g., pesticide registrants and permittees) to generate additional data to avoid the AFs and overly stringent criteria. This is not appropriate or an adequate scientific rational for deriving criteria based on insufficient data.

RESPONSE TO COMMENT 8-5: **UCD** See response to comment 6-15.

RB Staff did not suggest, nor is it staff's intent that a low data threshold would be misused to place inordinate burden for data generation on regulated entities. The Basin Plan already allows for interpretation of narrative criteria on the basis of only one data point through use of a static assessment factor ($1/10^{th}$ of the lowest LC50). In contrast, the proposed methodology provides for an assessment factor that decreases as the amount of data increases, and provides for a transition to a species sensitivity distribution method at an earlier stage than can currently be supported using the EPA methodology.

Nor is it inappropriate to establish methods that encourage rather than discourage generation of additional data. It is common practice in regulatory and non regulatory settings to transition to more sophisticated tools as needs, resources and available data increase. The choice of methodology approach as it relates to encouraging or discouraging generation of additional data is discussed in detail in Section 7.3.5 of the Phase 1 report. See also responses to comments 1-1, 6-15 and 6-16 for additional discussion about data availability.

COMMENT 8-6: The proposed procedure allows use of the Species Sensitivity Distribution (SSD) method to generate criteria for pesticide data sets with as few as 5 species. This is simply not enough data to adequately characterize a distribution. See also previous comment.

RESPONSE TO COMMENT 8-6: **UCD** See response to comment 6-7.

COMMENT 8-7: The proposed procedure uses the Burr III type distribution for the SSD and uses the entire distribution to estimate the 5th percentile (instead of focusing on the most sensitive species). In contrast to the U.S. EPA method, this allows greater influence on the criterion by relatively insensitive species, and requires the data set to conform to the underlying distribution assumption. The Burr III distribution's behavior should be evaluated to determine its performance with small data sets, and the potential effects on criteria of "outliers" and censored data. Although the Burr III family of distributions is fairly robust, this assumption requires a unimodal distribution without outliers to correctly estimate the 5th percentile. The derivation method would be improved by focusing on the sensitive end of the sensitivity distribution.

RESPONSE TO COMMENT 8-7: **UCD** While focusing on the sensitive end may have some advantages, it is preferable to use all the data, or split data based on modalities (if the distribution of data points indicates bi-modal distribution). There is less need to focus on the sensitive end if the model fits the data. BurrIII is much more flexible distribution, which makes is easier to fit different data sets. In Ch2 the fit of the BurrIII was evaluated and the BurIII was found to work as well as the other distributions or better for most data sets. The Australia/New Zealand scientists also did a lot of work validating the use of the Burr III family and found that outliers are usually not a problem.

Additionally, Chapter 2-3.1.5 discusses additional reasons to consider approaches other than U.S. EPA:

"…. there is no biological basis for selecting a triangular distribution, that not all of the data are used to fit the distribution, and that it assumes that a threshold toxicity value exists. …the U.S. EPA (1985) methodology derives criteria with no associated confidence levels. Thus, there is good reason to consider adopting a different SSD method for the new methodology. "

COMMENT 8-8: In responses to Regional Board staff comments, the authors state. . "The U.S. EPA method works reasonably well despite violations of distributional assumptions because the method ultimately focuses on just the four values nearest the 5th percentile, thus often disregarding a large body of available data." This is not a correct characterization. The U.S. EPA method uses all of the data in the distribution to establish the probabilities of the distribution. It also recognizes that results for insensitive species have little relevance and should have little influence on estimating criteria to protect sensitive

species. U.S. EPA's method has the advantages of making few assumptions about the underlying distribution and thus avoids potential problems of multi-modality and outliers in the data set. In many ways, it is a more flexible and robust method using the Burr III distribution, and should be reconsidered.

RESPONSE TO COMMENT 8-8: **UCD** In the EPA (1985) methodology, the actual values of any data above the lowest 4 data are not used directly in the calculation of the FAV. However, section 7.2.2.1 of the Phase I report more clearly discusses how these values beyond the lowest 4 are used to in establishing percentile ranks, specifically, they are used to calculate the total n in the rank calculation.

The log –triangular distribution is not without assumptions, as discussed in the response to comment 8-7. While no distribution may be perfect, the distribution chosen fits all the data better and avoids some of the mentioned problems. Also it is preferable to split the data if modalities are evident.

COMMENT 8-9: The proposed method treats potential outliers and bimodal distributions inappropriately by excluding data from the dataset without accounting for them in the probability distribution. This results in unnecessary reduction of the size of the data set, and consequently increases extrapolation and uncertainty in estimating the desired level of protection. This will result in unnecessarily stringent criteria. This deficiency in the proposed method can be addressed by the U.S. EPA method of estimating the 5" percentile value using only the sensitive end of the toxicity distribution. This method can be combined with the Burr III distribution fitting method.

RESPONSE TO COMMENT 8-9: **UCD** Outliers are no longer excluded. See response to comment 6-13.

COMMENT 8-10: The SSD procedure should also be refined to allow inclusion or consideration of results that may not meet all of the current data evaluation quality criteria. These data often provide enough information to include the result in the overall distribution of toxicity results. The current evaluation process is too quick to exclude results that can contribute to the overall distribution. Exclusion of useful results without adjusting the probability distribution will lower the criterion unnecessarily. This deficiency in the method should be addressed by modifying the SSD procedure to accommodate censored data (e.g., an indeterminate SMAV that is greater than a specific concentration) that doesn't overly influence the lower (sensitive) end of the distribution. This can be accomplished

through use of well-established "Regression on Order Statistics" (ROS) statistical estimation methods.

RESPONSE TO COMMENT 8-10: **UCD** Censored data (values reported as > or <) should be excluded because the actual effect or no effect levels could be many times the reported numeric value, and the distribution analysis cannot handle such data. Incorporating them could be done in some circumstances, but this is very dependent on the reason for reporting them this way and how much difference different possibilities of the true value would make. The authors have consulted with statisticians on this issue and have not found a method that would be able to easily analyze all of this data. Log-triangular works for high values only; ROS works for low values only (see below). Since their proper incorporation would be subjective with most distributions and their usefulness is limited, they will not be used in the distribution, but they will be considered as supplemental information.

Regression on Order Statistics (ROS) is not very appropriate to toxicity data that has a wide range of left and right censored data. ROS has only been used for left censored data, so it would not be able to handle all the types of censored data generated by toxicity studies. Also it has only been used for distributions that can be transformed to a normal, such as the lognormal. This would require that this methodology use a different distribution. The decision of which distribution to use should not be base on minority of data, which may be of questionable quality. Reasons for censored data may be well justified, such as toxicity occurring below detection limits or above the water solubility of the compound, but sometimes a censored data may simply be the result of appropriate concentrations not being used.

Further, while censored data are not used in the criteria calculation, if the study is rated as high quality otherwise the values will be included in the supplemental data set, which is reviewed before the criterion is set. In this way, if there is a high quality censored data that indicates toxicity may occur at a concentration lower than the calculated criterion the criterion may be adjusted downward to protect this species.

COMMENT 8-11: The use and basis of a default ACR of 12.4 when there are insufficient pesticide-specific ACR data is not valid. The basis for using an 80th percentile default ACR value is not adequately supported, and in the Phase I report TenBrook & Tjeerdema (2006) even concluded that there is no evidence that any default ACR value is appropriate for

pesticides. They subsequently offer the rationale that an ACR is needed to calculate a chronic criterion and that " ... The 80th percentile of values is used in the Great Lakes methodology (U.S. EPA 2003) and that is why it was selected for the new methodology." What is the underlying reason for using an 80th percentile value? If there are insufficient data to generate a valid pesticide-specific ACR, sufficient data should be generated instead of using a simplistic and scientifically invalid default value.

RESPONSE TO COMMENT 8-11: **UCD** Because there was no existing ACR that would be appropriate as a default ACR for pesticides (because existing ACRs were derived on many types of compounds), the new method derived a default ACR using all available ACRs from exiting pesticide criteria documents, as detailed in section 2-3.2.5.3. The procedure for deriving this factor was based on an extensive report by Host et al. (1995) in which they described both empirical and theoretical methods for derivation of factors using data sets for all kinds of chemicals. The 80th percentile was calculated in that report; however the decision to use it was from the Great Lakes Initiative. More data would produce more accurate and more pesticide-specific estimates. In the absence of such data, the chosen approach is the best available approximation.

2.9. Comment Letter 9 – Allen Short, San Joaquin Tributary Association

COMMENT 9-1: Overall, the proposed methodology appears to be sound and well documented. The authors provided a thorough response to previous Regional Board and peer review comments, and revised the approach and documentation when deemed appropriate.

RESPONSE TO COMMENT 9-1: **UCD/RB** Comment acknowledged.

COMMENT 9-2: The main concern we have with the methodology outlined in this document is use of an alternate approach (the “assessment factor method”) to develop criteria when limited data are available. Theoretically, this could result in adoption of a criterion when only one study result (for one species) is available. Due to the high level of uncertainty associated with this approach, it could either lead to significant underprotection or overprotection of the aquatic life community. Since the "assessment factors" increase with uncertainty, this method may result in an extremely low criterion value that is impractical and unwarranted. We question whether it is productive to attempt to develop

criteria for those pesticides with limited relevant, high quality toxicity data. Adopting any criterion with only one study result is not scientifically sound and will likely have no validity in the scientific community. While it may be appropriate to identify a screening threshold when only limited data are available, it would be misleading to attempt to develop a criterion that is supposed to be protective of the aquatic life community. In such cases, the effort might be better spent in conducting site-specific toxicity tests with resident species. In any case, the existing narrative toxicity objective in the Basin Plan could be used until sufficient data are available to develop a well-documented criterion.

RESPONSE TO COMMENT 9-2: **UCD/RB** See response to comment 6-15 and 8-5.

COMMENT 9-3: The proposed methodology includes an allowable frequency of exceedance of once every 3 years for both the acute and chronic objectives. This is the same as the frequency of exceedance allowed for in the U.S. EPA 1985 guidance, and the authors of the new methodology present good rationale for using this approach, given the limited supporting data available on ecosystem recovery after pulse disturbances. However, rather than requiring this frequency of exceedance to be applied in all cases, it would be reasonable to allow flexibility in the frequency of exceedance on a case-by-case basis, if sufficient data are available for specific pesticides. Although adequate relevant data are unlikely to be available in the near future, the proposed methodology should allow for consideration of new data that become available. There should also be flexibility to specify how exceedances could occur, in terms of duration and magnitude of exceedance. During the workshop on April 18, 2007, Regional Board staff stated that the guidance did allow for flexibility in the frequency of exceedance if data becomes available. Section 3-7.0 states that “These averaging periods may be modified if data and/or models become available that can scientifically defend altering them”. However, as currently written it is not clear whether the frequency of exceedance period could be modified – this should be made explicit.

RESPONSE TO COMMENT 9-3: **UCD** Section 3-9.0 (previously section 3-7.0) has been modified to allow flexibility in the averaging period and the frequency of exceedance. The Phase II Report also provides some discussion on what kinds of data might be useful. See Chapter 2-3.4

COMMENT 9-4: Section 2-3.6.2 (Bioaccumulation/secondary poisoning, p. 2-73) contains contradictory statements. The first sentence

states “This methodology is concerned with setting water quality criteria for the protection of aquatic life, thus it is not directly concerned with the protection of terrestrial wildlife or human health.” The third paragraph states that if a chemical is determined to have bioaccumulative potential, and dietary or FDA action levels are available, an additional analysis must be conducted and the criteria may need to be adjusted to ensure protection of wildlife/human health.

While it is important to consider bioaccumulative effects for protection of both wildlife and human health, this should be done as a separate and parallel process and a separate set of values should be developed based on protection of these receptors. Keeping the process separate will have multiple benefits, including:

· Better transparency in how criteria are developed; it will be clear which values are protective of aquatic life, which are protective of wildlife, and which are protective of human health. The lowest values can be applied as appropriate.
· Ease of use by risk assessors to determine which receptor groups are most at risk at specific locations, without having to go back to the derivation of the values
· Ease of use in determining which values are applicable to specific water bodies, based on beneficial uses

Many bioaccumulative compounds have been demonstrated to enter the food web primarily through a benthic (sediment pathway) rather than through a water column pathway. For those pesticides that are hydrophobic and likely to reside primarily in the sediments, food web effects might be better addressed through sediment quality objectives, which the Regional Board is also planning to develop. Note also that the proposed criteria are based on toxicity where the only mode of accumulation is via direct absorption from water (no dietary intake). This is not consistent with the food web effects that involve dietary uptake. Therefore, the proposed criteria should address “direct” toxicity to aquatic organisms, and not the “indirect” toxic effects due to food web biomagnification.

RESPONSE TO COMMENT 9-4: **UCD** This methodology is for water quality criteria only. This section (Bioaccumulation/secondary poisoning, now section 2-6.1,) is included to make sure this criterion is not in conflict with other protection goals, with the rationale that terrestrial wildlife/humans are drinking the water or eating aquatic organisms. The methodology had been clarified in this regard. A new section has been created in Chapter 2 with better instruction to help differentiate considerations that can be used to change the criteria (2-5.0) from those considerations that cannot be used to alter criteria (listed now in section 2-

6.0), such as the bioaccumulation section. A similar rearrangement has been made in Chapter 3.

RB For many pesticides that could potentially impact surface waters, aquatic life uses are the most sensitive beneficial uses. In defining the scope of work for the Central Valley Basin Plan Amendment, we assumed that protection of aquatic life beneficial uses will be protective of other beneficial uses of water. The bioaccumulation section simply provides one means for checking this assumption. The purpose of Section 2-6.1 is to ensure that in the process of protecting aquatic life, that a criteria is not established which could foreseeably have an adverse impact on wildlife and human health. The method is not intended to provide guidance on how an adjustment should be made if human health or wildlife beneficial uses are the most sensitive use. The methodology has been revised to clarify this point.

> **COMMENT 9-5**: Section 2-3.1.1: For the specific data sets analyzed in the report, Burr Type III distribution appears to be a good choice. However, one should not automatically assume this distribution for new data sets. Based on the results for the data sets in the report, one may define the null hypothesis to be that a new data set follows the Burr Type III distribution. That is, give the benefit of doubt to the Burr Type III distribution. But this hypothesis should be tested against the new data. If the data conclusively show that the null hypothesis should be rejected, then a search for a more appropriate distribution would be appropriate.

RESPONSE TO COMMENT 9-5: **UCD** BurrIII is a very rich family that contains or approximates a couple of the most commonly used distributions including the log-triangular and lognormal distributions. So in essence, it is doing as the commenter suggests. However, with five to eight points the Log-logistic distribution is to be tested first, then the BurrIII distribution (see response to comment 11-37, section 3-3.2.2). A fit test has been added to check that the distribution adequately fits the data (See section 3-3.2.4).

> **COMMENT 9-6**: The significance level (p value) for the Burr Type III distribution did not appear to be listed in any of the tables or figures. The p value for lognormal distribution is shown in Figure 2.1. Even when Table 2.3 shows that the fit number is better (lower) for Burr III than for lognormal, that does not necessarily mean that Burr III is a good choice. For example, Table 2.3 shows that, for Endrin, Burr III is better than Lognormal. However, Figure 2.1 shows the p value for lognormal to be 0.001. Thus, even though better than lognormal, Burr III may still be rejected based on p value.

RESPONSE TO COMMENT 9-6: **UCD** Later in Figure 2.2 and Table 2.3 another indicator of fit is shown for both distributions. BurrIII was chosen partly because it overall provided a better fit. There may be individual data sets that the BurrIII distribution may not fit. Accordingly, the method has been revised to include a fit test that generates a p value for BurrIII distribution, to avoid generating a criterion from a distribution that does not fit the data.

COMMENT 9-7: The standard approach proposed requires less toxicity data than the U.S. EPA approach (acceptable data on 5 taxa as compared to 8 taxa). However, the reasons they give for this do seem reasonable and justified (page 2-16). Our concern is related to the alternate "assessment factor" approach recommended when data on less than 5 taxa are available.

RESPONSE TO COMMENT 9-7: **UCD** See response to comment 6-15.

COMMENT 9-8: Table 3-15 specifies the assessment factors to apply when data sets include fewer than 5 values. What about the case when there are more than five values, but they do not cover the 5 types of organisms necessary to apply the SSD method (specified on page 3-8)? Theoretically, there could be 20 data points covering only 4 or less of the 5 organism categories – what is to be done in that case?

RESPONSE TO COMMENT 9-8: **UCD** The method for fewer than 5 data points would apply.

COMMENT 9-9: Values reported as > or < are excluded from the data set. Excluding censored data from a data set could introduce a significant bias in the results. There are statistical methods available to analyze censored data. Because the new proposed guidelines rule out more data than the U.S. EPA method, this can have a substantial effect on the value of the criteria derived, as documented in the case of chlorpyrifos.

RESPONSE TO COMMENT 9-9: **UCD** See response to comment 8-10.

2.10. Comment Letter 10 – Wendell Kido, Sacramento Regional County Sanitation District,

This comment letter is identical in content and language as Comment Letter 8 and is responded to above.

2.11. Comment Letter 11 – Lenwood Hall, University of Maryland,

COMMENT 11-1: A clear statement of specific goals is needed in the Introduction. The current text states that the goal is to "develop a methodology for derivation of pesticide water quality criteria for the protection of aquatic life in the Sacramento and San Joaquin River Basins". The critical point in this goal is what level of protection does the new method seek, i.e. protection of all species, 95% of the species as outlined in the U.S. EPA water quality criteria document (Stephen et al. 1985) or some other level of protection. U.S. EPA assumes that aquatic ecosystems can tolerate some stress and occasional adverse effects; therefore, protection of all species at all times and places is not necessary (Stephen et al., 1985).

RESPONSE TO COMMENT 11-1: **UCD** See response to comment 2-5.

COMMENT 11-2: It is unclear if the new methodology would apply to pesticides such as copper (a trace metal). If so, then the water quality effects section would need to be expanded to address water quality effects (i.e., hardness and dissolved organic carbon influence copper toxicity).

RESPONSE TO COMMENT 11-2: **UCD** Water quality effects are included in Section 3-5.0. Dissolved organic carbon is addressed more in Section 3-5.1 'Bioavalibility'. Most of this method can be used for metals. However, the default ACR and assessment factors were based on data from organics, so those particular procedures would probably not be most appropriate. Those sections (3-3.3 and 3-4.2.3) of Chapter 3 have been revised to state this. Some professional judgment will be needed if the compound is not a typical organic pesticide.

RB The method is most appropriate for use in deriving criteria for organic pesticides as the information used to develop the methodology were all form organic pesticides.. On a practical basis, water standards for the protection of

aquatic life for copper and other metals have already been established either in the Basin Plan or through the California Toxics Rule, so there is no need to derive criteria for most metals of concern. Staff has no plans to revisit the established metals water quality standards. .

COMMENT 11-3: The Introduction should also state why the Central Valley Regional Water Quality Control Board (CVRWQCB) has decided that a new criteria derivation method is needed. Does the CVRWQCB believe the existing criteria development methods used by U.S. EPA and California Department of Fish and Game (CDFG) are inadequate or in some way flawed?

RESPONSE TO COMMENT 11-3: **RB** See response to comment 1-1.

COMMENT 11-4: The new criteria development methodology should be "data driven" and require at least as much toxicity data as the U.S. EPA method (Stephen et al., 1985) to avoid uncertainty in the final acute and chronic criteria for pesticides. However, this is not the case as data for only 5 species are required with this new methodology for an SSD compared to 8 species required by U.S. EPA (Stephen et al. 1985). The use of 5 toxicity data points is problematic as Wheeler et al. (2002) states that 10 toxicity data points from individual species are needed for a reasonable SSD.

RESPONSE TO COMMENT 11-4: **UCD** Wheeler et al. was just one of several references disused in the Phase I report. Most of the rest of that discussion supported the choice of using 5 data for the SSD. See also response to comment 6-7.

RB See response to comment 1-1.

COMMENT 11-5: An even more troubling component of the new methodology is the use of assessment factors (also called safety factors, application factors and extrapolation factors) for pesticides with small data sets (less than five toxicity values for designated species). The use of assessment factors greatly increases the possibility of overestimating risk as reported by Chapman et al. (1998) and discussed by the authors. For example, the authors provide an example of how conservative the assessment factor approach can be with the chlorpyrifos example in Chapter 4. The final acute criterion derived by the new criteria method is 11.5 ng/L based on five acute data points but if only one data point had been available for Daphnia, the assessment factor approach would have

derived an acute criterion on 0.03 ng/L. This is an extremely low value, below 1 ng/L, that cannot be measured with current analytical methods.

RESPONSE TO COMMENT 11-5: **UCD** See response to comment and 6-15. Current detection limits should not be a driver of criteria derivation.

RB See response to comments 1-1, 1-4 and 6-15.

COMMENT 11-6: An example of unnecessary data reduction in the proposed methodology is the use of only North American species for criteria development. Given that the presence of limited toxicity data is a major issue with criteria development it would seem prudent to use a phylogenetic rather than geographic considerations when selecting toxicity data. If toxicity data were available for a non native North American species that has a closely related species present in North America then these data should be used for criteria development if the study is acceptable based on the data screening process. This approach would be acceptable since Suter 1993 has demonstrated that closely related species have similar sensitivity to contaminants. As stated by the authors, the best way to minimize overprotection and provide science based criteria is to expand available acute and chronic toxicity data sets. I would strongly support this recommendation and promote a science based "data driven" approach.

RESPONSE TO COMMENT 11-6: **UCD** The method includes all species from taxonomic families represented in North America, so included are species in North America plus related species that do not reside in North America but have related relatives in the same taxonomic family that do reside in North America.

COMMENT 11-7: There is a continual theme throughout the report that various critical components of the new methodology are policy decisions, i.e. acceptance of certain toxicity data, selection of certainty levels in tails of species sensitivity distributions, and determination of assessment factors. I strongly disagree with this approach because empirical science should be used to determine the various critical components of the new criteria methodology. Both qualified scientists and policy types should work together to develop the various components of the new criteria methodology.

RESPONSE TO COMMENT 11-7: **UCD** The acceptance of certain toxicity data, selection of certainty levels in tails of species sensitivity distributions, and determination of assessment factors were not policy driven. The decision to

generate criteria in general is policy driven. However, the procedures are not policy driven. The selection criteria are stringent, but clearly scientifically based. It is agreed that derivation of criteria benefits from a large data set. However, the reality is that such data sets are simply not available for a large number of pesticides. No other better ways were found to choosing a percentile cut off or to handle small data sets.

RB To the extent feasible, the Regional Board has endeavored to separate policy issues from science issues. The main method work has been intentionally contracted out to UC Davis instead of being developed in house in part to help insulate the method from policy issues. The methodology has been submitted to scientific peer review by a panel of experts outside of the Regional Board. All of the substantive comments about the method are included in the Peer Review Addendum, which is available at the Boards Website at:

http://www.waterboards.ca.gov/centralvalley/water_issues/tmdl/central_valley_projects/central_valley_pesticides/criteria_method/index.shtml

A review of those comments shows that substantive comments by Regional Board Staff were specifically withheld from UC Davis until after Peer review comments had been received. With that understanding, there are some legitimate policy concerns that must be addressed by the method. These are described in the project scope of work, which is available at the above website address. The primary policy concern is that the method must generate criteria that meet the legal mandate to be protective of aquatic life beneficial uses. In some places, the method recognizes that legal requirement by providing for calculation of different percentile cutoffs. The scope of work also requires that the method be able to accommodate data sets of varying size. This reflects the reality of how much data is available on pesticides. See also responses to comments 1-1 and 3-1.

> **COMMENT 11-8**: For acute criteria, a 1-h averaging is proposed while for chronic criteria a 4-d period is established. These two averaging periods are used by U.S. EPA in their criteria development method (Stephen et al., 1985). It is important to remember that the U.S. EPA approach developed in the mid 1980s was primarily developed for POINT SOURCE discharges where constituents such as ammonia are measured at frequent intervals (hourly or daily). However, for pesticides hourly measurements are rare for monitoring effects in California. Daily measurements for four consecutive days are somewhat more likely but are still the exception and not the rule for pesticide monitoring studies in the Central Valley. Therefore, the basis for using 1-h (acute criterion) and

4-d (chronic criterion) averaging periods for allowable exposure duration for pesticides in the Central Valley is not appropriate. Pesticide data collected from monitoring studies in the Central Valley and obtained from California's Department of Pesticide Regulation should be reviewed to determine the most common frequency of pesticide measurements (i.e., once a month for a year) and these data should be used to select the most appropriate averaging periods for both acute and chronic criteria.

RESPONSE TO COMMENT 11-8: **UCD** The averaging periods are based on potential biological effects and observations of organism recovery. The length of time needed to avoid averaging out high concentration pulses that may be high enough or long enough to be toxic is considered. They aim to be protective in light of a generally high level of uncertainty, especially for chronic effects. The review in Phase II considered pesticide applications or run-off resembling brief and mild exceedance, not point source discharges.

RB The proposed methodology is consistent with existing Basin Plan monitoring requirements. Past pesticide water quality criteria were derived using the U.S. EPA method and current implementation plans reflect that including the 1-hr and 4-day averaging periods.

COMMENT 11-9: In setting an allowable frequency of exceedance of acute and chronic criterion, the key question is how much time is needed for organisms at various levels of organization to recover from brief pulse exposures to contaminants. The proposed criteria method recommends an allowable frequency of exceedance of once in three years. This is the same frequency of exceedance used by the U.S. EPA in their criteria method (Stephen et al., 1985) and as stated by the authors the 3-year frequency of exceedance was supported by minimal data. The receptor group (most sensitive biological assemblage) for any given pesticide should be considered when establishing the frequency of exceedance for a specific type of pesticide. For example, the receptor group for herbicides is plants such as phytoplankton which have short life histories (several days). Therefore, a once in three years exceedance is overprotective for species such as phytoplankton which can recover within days or weeks. In contrast, for species with long life cycles (greater than 5 years) such as various fish, a once in three year exceedance may be appropriate. I would recommend flexibility for the frequency of exceedance component of the new criteria development method that would allow the use of life histories for receptor species in order to determine the most appropriate frequency of exceedance.

RESPONSE TO COMMENT 11-9: **UCD** The basis for establishing a once in three year frequency is thoroughly discussed in Section 2-3.4.1 of the Phase II report. As discussed in this section, predicting the outcome of an exceedance varies on many factors. With all these factors to consider, it would be difficult to determine allowable exceedances based on a particular chemical or the effects of one compound on a particular group of organisms. This would be dependent on the information available for that particular compound. Additionally, toxicity to short lived species can have a ripple affect up the food chain. Since there is a limited amount of information available to assess such specific questions for many chemicals, the more general question of what is the maximum time needed to recover for any probable exceedance can be better answered with the information currently available.

> **COMMENT 11-10**: The authors should also explore the use of the binomial approach for determining the number of pesticide exceedances needed before a violation occurs. The California State Board uses the binomial approach for listing and delisting impaired water bodies in the State based on exceedences of both toxicants (i.e. pesticides) and conventional pollutants (i.e., pH, dissolved oxygen) (SWRCB, 2004). The binomial approach has statistical underpinnings that allows the determination of error rates associated with impairment declarations and a process to limit error rates.

RESPONSE TO COMMENT 11-10: **RB** The Binomial Method is a statistical test used in the development of the 303(d) list under the State's Listing Policy. As stated in its introduction, the Listing Policy is not intended to establish water quality criteria (SWRCB 2004).

> **COMMENT 11-11**: Using a species sensitivity distribution (SSD) for criteria derivation requires selection of a percentile of the distribution as a cutoff point. An interpretation of this cutoff point means that species lying above this point in the distribution will be protected as long as the concentration of the chemical is below the concentration at the selected percentile. The authors state that species lying below percentile would be harmed. This is incorrect. Species lying below this percentile would not be fully protected but not necessarily harmed. The authors state that the choice of the 5th percentile is purely pragmatic and has been used by other organizations such as U.S. EPA (Stephen et al., 1985), the Dutch (RIVM, 2001), and Australia/New Zealand (ANZECC and ARMCANZ, 2000) without rigorous scientific justification. Therefore, scientific rationale should be provided before the 5th percentile is used in the new criteria

development method. In addition, scientific rationale should be provided to justify dividing the 5th percentile by a factor of 2 before determining the final acute value.

RESPONSE TO COMMENT 11-11: **UCD** The discussion in the new method agrees with the commenter's interpretation of the 5th percentile. Chapter 2-3.1.2 includes a discussion of the common misperception that species below the 5th percentile would necessarily be harmed. Section 2-3.1.2 states that, "Van Straalen & Van Leeuwen (2002) note that it is *not correct* to interpret the 5th percentile to mean that 5% of species will be harmed (as was argued, for example, by Lillebo *et al.* 1988, regarding the U.S. EPA 1985 methodology). Rather, this approach is one method for derivation of a predicted no-effect concentration, and although the choice of the 5th percentile is a purely a pragmatic one, it *has been validated by field studies*." (emphasis added). Section 2-3.1.2 also contains a review of studies that found agreement between the 5th percentile and the NOEC

The safety factor of 2 is applied because the SSD is constructed with toxicity values that indicate a 50% effect level. This figure was based on 219 acute toxicity tests with various chemicals, which showed that the mean concentration that did not cause mortality greater than control was 0.44 times the LC50 (34 FR 97, p 21508-21218). The inverse of .44 (2.27) was rounded to 2 for use in EPA methods. The method has been clarified to include this information.

COMMENT 11-12: As stated by the authors, the final element to consider is whether a pesticide that is present in water at a criterion level might have the potential to move from that water compartment into another environmental compartment (i.e., sediment, biota, air). This harmonization issue will be particularly important for hydrophobic pesticides, such as pyrethroids, that may eventually concentrate in bed sediment. Therefore, water quality criteria and sediment criteria for pesticides such as pyrethroids need harmonization to avoid possible conflicts. This would involve communication between Regional Board scientists and State Board scientists (i.e., Chris Beegan) that are addressing these water quality and sediment quality criteria issues.

RESPONSE TO COMMENT 11-12: **UCD** This is already included in Section 3-7.2 Harmonization/coherence across media. The method has been revised to clarify that if conflict is found, additional review may be required. Also see response to comment 9-4.

RB The proposed method is intended to establish criteria specific to the water column. As discussed by UC Davis, the method includes a means to evaluate harmonization of the criteria across media to reveal when a criterion may not be protective to other media, such as the sediment compartment. However, establishment of sediment quality criteria or criteria within any other compartment is beyond the scope of this project. If the method indicates that a water quality criterion may not be protective of aquatic life within other environmental compartments, additional evaluation will be required.

> **COMMENT 11-13**: When will the Phase III report be available? What process will follow the Phase III report? Will a Basin Plan amendment be developed to approve the new criteria methodology? Will documentation be provided on how the Regional Board has responded to review comments from the public/interested parties?

RESPONSE TO COMMENT 11-13: **RB** Data gathering for phase III has already begun. Derivation work will be conducted once Phase II (method development) is completed, estimated September 2009. UC Davis will finalize one or more "Phase III" pesticide criteria reports following scientific peer review. Regional Board Staff will perform additional review of the resulting criteria, along with criteria from other sources, as part of the water quality objectives alternatives evaluation that will be part of a Central Valley Pesticide Basin Plan Amendment Staff Report. The first Basin Plan Amendment Staff Report is scheduled to be submitted to the Public for review in early 2010 and brought before the Regional Board by June 2010. Subsequent amendments for other pesticides will likely following in the next year. The methodology is a technical report produced by UC Davis, it is not intended as a policy or regulation for Regional Board adoption, but Criteria derived using the methodology will be among those considered by the Board for establishing water quality objectives. Docuentation of responses to comments form the publicy, such as this document, will be provided in response to comments received during comment periods. Written comments will be responded to in accordance with standard Board policy and procedures.

> **COMMENT 11-14**: I would suggest rounding off the acute chlorpyrifos criterion to 12 ng/L and the chronic criterion to 11 ng/L to reflect the sensitivity of the analytical method for chlorpyrifos measurements. It is also prudent to check monitoring data for chlorpyrifos in the Central Valley to see if the current analytical detection limits for chlorpyrifos used by most laboratories are below or above the proposed criteria.

RESPONSE TO COMMENT 11-14: **UCD** The acute and chronic criteria for chlorpyrifos has been rounded to one significant figure in the revised report,,

RB The scope of work for the method is to derive criteria that should be protective of aquatic life beneficial uses. Feasibility issues, such as the ability to detect a pesticide at a given concentration are considered once the criteria are proposed adoption as a water quality objective. Also see response to comment 1-4.

> **COMMENT 11-15**: The last sentence in paragraph 1 of Page 2-1 states that 11 other pesticide data sets were used from EPA; however, only 9 references are provided.

RESPONSE TO COMMENT 11-15: **UCD** One of the references is for both Dieldrin and Aldrin. One reference for DDT seems to have been mistakenly left out. The report has been revised to include the following reference:

U.S. EPA. 1980g. Ambient Water Quality Criteria for DDT, EPA 440/5-80-038. United States Environmental Protection Agency, Washington D. C.

> **COMMENT 11-16**: The list of acute methodologies on page 2-2 should include methodologies for plants.

RESPONSE TO COMMENT 11-16: **UCD** Life cycles of plants vary widely and procedures for conducting toxicity tests with plants are not well developed. Currently plant toxicity tests usually measure endpoints generally associated with chronic toxicity, such as growth and reproduction. Therefore, the methodology has been revised, removing the plant requirement for the acute distribution, but leaving it in the chronic distribution. Also an option is provided for the use of the lowest plant NOEC in case of limited data (similar to EPA methods) see section 3-4.3.

> **COMMENT 11-17**: It is unclear how the use of non-traditional endpoints may be used to derive criteria if those endpoints have been adequately linked to effects on survival, growth and reproduction or population parameters. Who makes this very critical decision on the use of non-traditional endpoints for criteria derivation? (a panel of experts, Regional Board scientists).

RESPONSE TO COMMENT 11-17: **UCD** Only non-traditional endpoints that have been linked to survival growth or reproduction are to be used in the calculation. The decision of whether or not there is a good link would primarily come from discussions in the peer reviewed literature. The fact that it is peer reviewed helps to create an objective discussion about the significance of the endpoints and whether or not an endpoint can be linked to a population effects. The ultimate use of the information and its accurate reporting will depend on the scientist working on the criteria report. The report describes this scenario for AChE inhibition (3-2.1.1.3) for which there is a decent amount of literature supporting the link to survival. At the very least, such studies can be included as supplemental data and the criteria discussed in this context.

RB The use of non-typical endpoints will be considered by the person running the method and vetted through the peer review/public review process. Ultimately, the Board will decide what criteria will be used for adoption as objectives.

> **COMMENT 11-18**: Microcosm and mesocosm data should used in the criteria derivation process if it is available and valid. For example, if microcosm/mesocosm NOECs/LOECs are substantially higher that the acute or chronic criterion then the data used to develop the criteria should be reevaluated.

RESPONSE TO COMMENT 11-18: **UCD** See response to comment 4-3.

> **COMMENT 11-19**: The rationale behind using the 75th percentile of scores for the reliability rating is needed. Chlorpyrifos may not be a good data set to use for this benchmark since this is a fairly rich data set and most other pesticide toxicity data sets will be less extensive.

RESPONSE TO COMMENT 11-19: **UCD** See response to comment 5-5.

> **COMMENT 11-20**: The point concerning considerable variability of sensitivity (on page 2-17, paragraph 4) between species within a genus is generally not supported by most of the literature. Suter (1993) that showed similar species have similar responses to chemicals.

RESPONSE TO COMMENT 11-20: **UCD** The protection goal of the proposed method is at the species level so it follows that species mean values will be used. Chapter 2-2.7 outlines the information found in the literature and how the case

was not strong enough to change to using the genus mean values instead of species mean values for the criteria calculation.

Additionally examples have been found in chlorpyrifos and diazinon criteria reports that do not agree with Suter (1993). The Chlorpyrifos criteria report that is part of the proposed method has species mean values for *Daphnia* that have a fairly wide range: 0.03 µg/mL (*ambigua*), 0.25 µg/mL (*pulex*), and 1.0 µg/mL (*magna*). Also the EPA criteria for Diazinon has species mean values for two species in the genus *Onchorhynchus:* 426 µg/mL (*mykiss*) and 2166 µg/mL (*clarki*), which differ almost by a factor of 5.

> **COMMENT 11-21**: Why were the 12 pesticide listed in Table 2.1 selected as test cases for the SSD method? Do they cover all the classes of pesticides (i.e. organophosphates, herbicides) that are used in the Central valley?

RESPONSE TO COMMENT 11-21: **UCD** They do not cover all the pesticide classes used in the Central Valley. These were used because they were all the available existing pesticides data sets from past EPA criteria. Collecting data and assembling new data sets is an enormous amount of work, so existing data sets were used as test cases for the SSDs.

> **COMMENT 11-22**: Rather than prescribe the distribution to use for the pesticide toxicity data, Burr III distribution, why not use the distribution that best fits the data?

RESPONSE TO COMMENT 11-22: **UCD** See response to comment 9-5.

> **COMMENT 11-23**: The points made by Chapman et al. (1998) and discussed by the authors on pages 2-47 and 2-48 would seem to justify why “Assessment Factors” should not be used to establish criteria, i.e. it greatly increases the possibility of overestimating risk

RESPONSE TO COMMENT 11-23: **UCD** See response to comment 6-15.

RB See response to comment 1-1 and 6-15.

> **COMMENT 11-24**: The use of toxicity data from the daphnia family for limited toxicity data sets may be overprotective for OP insecticides, since

the receptor and most sensitive taxa are cladocerans. However, the use of daphnids may be under-protective or inappropriate for other pesticides where daphnids are not the receptor taxa.

RESPONSE TO COMMENT 11-24: **UCD** This is why data from other groups are required for the SSD. The requirement of other species for the SSD should prevent the criteria from being under protective. And an SSD with daphnid data should protect daphnids, which is desirable, and not likely to be overprotective.

But it is assumed that the commenter is referring to the assessment factor method and not the case with enough data to do an SSD. Generally daphnia are the most sensitive so that was why they are required for all calculations. This is the first part to avoid under-protection. Also, if the occasion arises when only the daphnid requirement is met and other species are lacking, there is a large safety factor to prevent the assessment factor method from being under-protective.

With limited data over protection is a possibility of the assessment factor method, but in an effort to meet the protection goals, some overprotection is preferred to underproduction to a reasonable extent.

COMMENT 11-25: Table 2.6 provides a clear example of why limited toxicity data (n<5) should never be used to establish criteria.

RESPONSE TO COMMENT 11-25: **UCD/RB** See response to comment 1-1 and 6-15.

COMMENT 11-26: Saltwater taxa should only be used if the pesticide toxicity data shows that salinity does not affect the toxicity of the pesticide. For example, salinity affects the toxicity of metals such as copper.

RESPONSE TO COMMENT 11-26: **UCD** In the new methodology, saltwater data is only used for the derivation of the acute-to-chronic ratio.

COMMENT 11-27: Why did the Great Lakes guidance document select the 80th percentile as a default value of ACRs? It should be stated clearly that if an ACR is available for a pesticide (as is the case for chlorpyrifos) then this ACR is used and not the default value of 12.4. The ACR for lindane is higher than the other ACRs in Table 2.8 and should therefore be checked carefully before including in this table

RESPONSE TO COMMENT 11-27: **UCD** It is clear in the method that data for specific pesticide is used first to calculate the ACR, and that if specific requirements are not met, then the default value may be used.

The ACR for lindane was from EPA criteria report on lindane (U.S. EPA 1980). This report included ACRs for Daphnia magna: 33, Chironomus tentanus: 63, and fathead minnow: 7.5. The criteria report also describes the physical and chemical properties of lindane and its toxicological effects. Lindane is very lipophilic and stable compared to other compounds in the table, except for chlordane, which it is similar to in these respects. These two compounds have the largest ACRs in the table. The mechanism of toxic action, as well as the chemical properties of contaminants determine their behavior and effects on organisms. Organochlorine pesticides generally are not as acutely toxic as other neurotoxic insecticides, but the chronic effects at very low levels are severe, in part due to their bioaccumulation potential. This explains the high ACR.

> **COMMENT 11-28**: It would be more reader friendly to include the frequency and duration components within the same section or subsection since these components of the criteria are closely tied together.

RESPONSE TO COMMENT 11-28: **UCD** These ideas are related. However, the current sequential arrangement doesn't seem to be generally confusing, and changing the document arrangement to include both in the same section would require significant changes to other parts of the document, so this has been left as it is.

> **COMMENT 11-29**: The comments that chlorpyrifos and diazinon are not fast acting toxicants is not supported by the newly derived chlorpyrifos acute and chronic values which are nearly identical and a previously published EPA diazinon criterion of 100 ng/L for both the acute and chronic criteria (U.S. EPA, 2000).

RESPONSE TO COMMENT 11-29: **UCD** The comment about not being fast acting was not in regard to chronic data. This comment was included in the discussion of acute averaging periods, to point out that toxicity may occur in the first 24 h of a 96 hr test, or in the case described, fairly evenly throughout the entire 96 hr period. The slower effects mentioned still occurred within the 96 hrs. Both methods are using 96 hr test data for acute data, since 'fast 'and 'slow' occurred within 96 hrs the relative speed of action will not make a difference in the chronic data, or in a comparison of acute vs. chronic criteria.

COMMENT 11-30: Semi-permeable membrane devices (SPMDs) are passive sampling devices that are intended to mimic uptake of bioavailable contaminants. The various negatives associated with using SPMDs are presented by the authors (i.e., they do not give quantitative results for polar organics). Therefore, I would not support the use of SPMDs for assessing bioavailability. However, if tissue data were available for resident bivalves in a particular study area potentially impacted by pesticides or well designed caging experiments with bivalves were conducted in the study areas, these data may be useful for addressing bioavailability issues. In order to address the issue of bioavailability, the new criteria method needs to have some flexibility to address this issue on a "pesticide specific basis" depending on the data available and the physical/chemical properties of the pesticide.

RESPONSE TO COMMENT 11-30: **UCD** Bioavailability information is discussed and considered in the criteria development and a few different method options are suggested. Caged bivalve studies are more appropriate to include in the bioaccumulation section. They involve uptake from feeding and are more species specific. The bioavailability section focuses on lower availability due to sorption to particles or solids.

COMMENT 11-31: The authors must be careful when evaluating possible additivity of chemicals with similar modes of action. The first consideration is that the chemicals must co-occur in the environment (present in the same sample). The next consideration is that additivity can not be assumed if measured concentrations of pesticides are below a certain threshold (Dr. Allan Felsot, personal communication, Washington State University).

RESPONSE TO COMMENT 11-31: **RB** The additivity issue to which the commenter refers, first appeared as a peer review comment (Felsot, 2005) on the San Joaquin River OP Pesticide Basin Plan Amendment (Beaulaurier et al., 2005). The peer reviewer stated that the Basin Plan's additivity formula, proposed for use in the San Joaquin Amendment did not reflect additive toxicity. The peer reviewer suggested an alternative method for calculating additive toxicity. Staff reviewed this comment and determined that the alternative method recommended by the peer reviewer is mathematically equivalent to the Basin Plan formula for additive toxic effects of pesticides (Beaulaurier et al., 2005, McClure et al., 2006).

In addition, staff noted that the purpose of the additivity formula is not to predict a given level at which impairment of beneficial uses might occur, but to identify a protective level below which no adverse effect would be expected, consistent with the legal mandate of the Board (Beaulaurier et al., 2005). The recommendation for the San Joaquin River Amendment, which was also adopted as part of the Delta Amendment, was to continue using the existing Basin Plan's additivity equation. Subsequent reviews of the Delta Amendment by Dr. Felsot (2006) concurred with the Board's decision.

No scientific evidence has been provided to support the suggestion that pesticides can be ignored at low levels when other pesticides having similar modes of action are present. In fact, studies by Deener et al. (1988) suggest that there is no such threshold for chemicals with a similar mode of action.

> **COMMENT 11-32**: It seems a stretch to include protection of terrestrial wildlife or human health within this report since the goal is to develop a criteria method for protection of aquatic life.

RESPONSE TO COMMENT 11-32: **RB** The scope of work was limited to protection of aquatic life because that is generally the most sensitive beneficial use where pesticide discharges are concerned. However, it is useful to verify that assumption. As such, it is appropriate that the methodology consider whether derived criteria will be protective of other beneficial uses, such as wildlife and human health. If the method indicates that a water quality criterion may not be protective of other beneficial uses, additional evaluation may be required. The method has been revised to clarify where additional review may be required. See also response to comment 9-4.

> **COMMENT 11-33**: It would seem appropriate to have qualified individuals from EPA, U. S. Fish and Wildlife Service and NOAA, who work with threatened and endangered species, review this part of the criteria development method.

RESPONSE TO COMMENT 11-33: **RB** US EPA has reviewed this document and provided comments (See Comment Letter 12 and 13). The California Department of Fish and Game, the state agency charged with protection of threatened and endangered species, were invited to act as peer review and participated in that capacity during Phase I of the project. The Phase II report was sent to CDFG, but no comments were received.

COMMENT 11-34: Definitions/references for plant toxicity testing should be included both the acute and chronic definition sections of chapter 3.

RESPONSE TO COMMENT 11-34: **UCD** See response to comment 11-16.

COMMENT 11-35: How does the toxicity data screening process developed by the authors compare with the data screening process used by U.S. EPA for their development of water quality criteria.

RESPONSE TO COMMENT 11-35: **UCD** Please see response to comment 4-2.

COMMENT 11-36: The authors support the use of a statistical test for outliers (Sokal and Rohlf, 1995) to delete suspect data points. Although this statistical approach is admirable, the authors may also want to consider the approach used by EPA for addressing outliers. EPA has addressed this issue in Stephen et al, (1995) by stating the following "acute values that appear to be questionable in comparison to other acute or chronic values for the same species or other species in the same genus should not be used in the calculation of species mean acute values".

RESPONSE TO COMMENT 11-36: **UCD** The outlier test has been removed and critical examination of data has been emphasized. See 6-13 for details on new procedure for outliers.

COMMENT 11-37: The authors state that the BurrliOZ software comes with a caution that for data sets of 8 or fewer values there is a great uncertainty in the calculated toxicity values. This provides further support for one of my main comments (see general comments section) that the use of 5 data points is too data restrictive and will produce SSDs and 5th centiles with a high degree of uncertainty.

RESPONSE TO COMMENT 11-37: **UCD** The software comes with a caution; it does not say the results are invalid. The Australia/New Zealand methodology also decided this software could be used with 5 data to set their target values. The software also comes with a procedure for 8 or fewer data of using the log-logistic distribution if it appears to fit better. This has been slightly revised to call for the log-logistic distribution to always be tried first. If log-logistic fails to fit the data then the BurrIII is to be used. This was done because the log-logistic distribution has one less parameter that will allow one more datum beyond minimum requirements to be used to establish the fit. For small data sets this

can help reduce variance that would accompany a distribution that had one additional fit parameter, like the BurrIII. Also a fit test has been added to ensure that one point isn't having an overwhelming influence in these smaller data sets.

> **COMMENT 11-38**: The first parameter listed in Table 3.7 is "results published or in signed, dated form". It is unclear to me what the term "published" means. If this means published in the "peer reviewed" literature then data reported from this type of reference should score higher than a published report that has not been subjected to "peer review".

RESPONSE TO COMMENT 11-38: **UCD** The idea is that there is some record of the data that can be found by others. It doesn't have to be peer reviewed.

> **COMMENT 11-39**: In developing Table 3.16 for ACRs that results in an ACR default value, I would suggest that ACRs for all classes of pesticides (i.e., herbicides, carbamates) that are suspected as potential stressors in the Central Valley be included in this table. This would provide a more representative ACR default value for the geographic area of concern.

RESPONSE TO COMMENT 11-39: **UCD** All ACRs derived by EPA methods for criteria are included in this section. There are no other multi-species ACRs to include. Using data only for pesticides used in the Central Valley would limit the section further. At this time it is not feasible to narrow it down by geographic region. The proposed method suggests incorporation of new ACRs as they are generated.

> **COMMENT 11-40**: The chronic value of 1 ng/L for Neomysis mercedis is extremely low and suspect. The study and ACE analysis that derived this value needs to be carefully reviewed. The authors later mention on page 4-6 that the Neomysis chronic value is not used to calculate the ACR. This point should be stated on page 4-5.

RESPONSE TO COMMENT 11-40: **UCD** This data was not used for deriving criteria. The report has been revised to reflect this more clearly.

> **COMMENT 11-41**: The authors need to provide more details in Section 4-6 on how the 5th and 1st percentiles were determined.

RESPONSE TO COMMENT 11-41: **UCD** This is done by using the software program referenced in the methodology. The corresponding section of chapter 3 (3-3.2.1) also includes more details about the procedure the software uses.

COMMENT 11-42: The authors state that Table 4.9 shows synergistic ratios. The Table 4.9 in my downloaded copy is Neomysis raw acute data from CDFG.

RESPONSE TO COMMENT 11-42: **UCD** Table 4.10 shows synergistic ratios. This has been corrected.

COMMENT 11-43: The authors explain the differences in their new lower acute and chronic criterion for chlorpyrifos compared with the EPA values or CDFG values by stating that different data sets were used for final calculations. A table or series of tables should be developed to clearly show why the chlorpyrifos criteria are different among the three methods, i.e., the new method, the EPA method and the CDFG method.

RESPONSE TO COMMENT 11-43: **UCD** Please see response to comment 4-2.

2.12. Comment Letter 12 – Joe Beaman, U.S. EPA

COMMENT 12-1: The authors' inclusion of comprehensive guidance for data evaluation and filtering is practical, and allows for objectivity and transparency in this important and sometimes controversial step of the criteria derivation process. This guidance may have widespread value in the near future, and the authors should share this guidance with other state and Regional EPA colleagues. Wide acceptance of these guidelines would go a long way in ensuring criteria derivation is based on sound scientific data, regardless of the entity performing the derivation.

The authors also provide a number of flow charts, data summary sheets, web addresses and tables are provided that will help provide guidance for determination of physical chemical parameters, default acute-to-chronic ratio, and other statistical tools and statistically-based tools and support values.

RESPONSE TO COMMENT 12-1: **UCD/RB** Comment acknowledged.

COMMENT 12-2: In Step12. v. d), the authors recommend removing statistical outliers, even if no scientifically acceptable reason for their deletion is present, they further recommend using "procedures" to determine if the potentially over or underproductive criteria as calculated provide adequate protection. In Section 3-6, they recognize that criteria may need to be adjusted "downward" based on numerous considerations such as ecosystem protection, presence of threatened and endangered species, etc., but provide no objective guidance on how this should be done. What type of value do the authors intend? A safety factor? Uncertainty factor? What will be its magnitude? What will it be based on?

RESPONSE TO COMMENT 12-2: **UCD** The outliers section was rewritten without the statistical test and to emphasize critical examination of data. See response to comment 6-13.

More guidance has been added from chapter 2 to section 3-6.0 on the downward adjustment. The recommended means of making such a downward adjustment is to use either a lower 95% confidence limit estimate of the 5th percentile (see discussion in section 2-3.1.3), or a median or 95% confidence limit estimate of the 1st percentile. Some judgment will be involved, but essentially a level that provides adequate protection in light of the information obtain in section 3-6.0 should be used.

COMMENT 12-3: Since the Central Valley Waterboard is developing a new methodology and exploring better ways to derive pesticide criteria, would it be more appropriate to derive the criteria based on which distribution fits the data best on a case-by-case basis, rather than relying on a single (albeit flexible) distribution?

RESPONSE TO COMMENT 12-3: **RB** Please see response to comment 9-5.

COMMENT 12-4: The 1985 Guidelines (U.S. EPA1985), require 8 specific taxa be present at a minimum to derive criteria, so as to ensure adequate protection for aquatic life. The proposed methodology removes requirements for the third chordate taxa (fish or amphibian), a non-chordate family such as mollusks and rotifers, and additional taxa not already represented in the dataset, while adding in a specific requirement for an alga/aquatic vascular plant when deriving criteria for herbicides. This approach is potentially problematic in that:

a) By inclusion of a plant in the data set, one might underestimate the toxicity of an herbicide if the remainder of the dataset were relatively insensitive. The 1985 Guidelines require a separate assessment for plant sensitivity by comparison of the Final Plant Value with the derived criteria, rather than implicit inclusion in the data set.

RESPONSE TO COMMENT 12-4: **UCD** The final criteria should be protective of all species mean values used in the SSD. If the 5th percentile is above the lowest known effect level, the methodology calls for adjustment bellow that, so that criteria are protective (to the best of our knowledge) of all species. If plants are the most sensitive species, a separate plant value may be calculated. For new method in the case that plants are most sensitive, please see response to comment 11-16.

RB The project scope of work includes the goal to be able to utilize datasets of varying quantity. If data are available for additional species, such as frogs or non-chordates, they will be used by the new methodology. Unfortunately, these species are not necessarily required for pesticide registration and are frequently unavailable. Under the current system, the lack of data would preclude use of the EPA methodology and criteria would be established based on 1/10th of the lowest LC50. In contrast, the proposed method allows consideration of all available data to derive criteria that are less likely to be over- nor under-protective. See response to comments 1-1 and 6-15 for additional discussion.

COMMENT 12-5: Removal of the third chordate may be potentially problematic, especially if amphibians are innately sensitive to the pesticide for which criteria derivation is being performed. A re-assessment of the proposed atrazine criteria is in progress due to concerns of amphibian toxicity. It is evident that at least frogs are resident to San Joaquin Valley (Moyle, 1973). If amphibians (or other required taxa – based on the 1985 Guidelines) are not present, the Central Valley Water Board should provide taxa lists for the watersheds in question to demonstrate that potentially sensitive species are not present.

RESPONSE TO COMMENT 12-5: **UCD** The new methodology will use data for all aquatic life species that are available. Amphibians are not specifically required by the EPA 1985 guidelines either, as stated in comment 12-4. Requirement "c" can be a fish or amphibian. If the commenter is referring to threatened and endangered species, a review of threatened and endangered species from the California Department and fish and Game lists is already included in section 3-6.3 of the method. Threatened and endangered species will

certainly be addressed as data are available. However, all taxa cannot be required in the new method (nor does EPA), so there may be an instance when a threatened or endangered spices is not represented in a criterion calculation.

In regards to the specific concern about atrazine, this comment possibly refers to developmental/endocrine effects of atrazine on amphibians; such effects can occur at very low concentrations. However, in general, amphibians are far less sensitive to pesticides than planktonic crustaceans or insect larvae.

RB See response to comments 1-1 and 12-4.

> **COMMENT 12-6**: Removal of the additional non-chordate requirement may also reduce protectiveness of criteria. In the case of ammonia, recent work has demonstrated that listed species of unionid mussels are particularly sensitive to the effects of ammonia. (Augsberger et.al. 2003). Also, as referenced in your methodology, chronic criteria for the pesticide, tributyltin, was based on endocrine effects on marine snails (U.S. EPA, 2003).

RESPONSE TO COMMENT 12-6: **UCD** See response to comment 12-4. To the extent it is available; all this information is included and considered in the new criteria derivation method, either in the SSD calculation, or in the supplemental info. It is just not required if the data is not available. If these data were required for all criteria, it would likely prevent setting criteria for a number of pesticides for which these are not the most sensitive species. It is probably very rare that mussels will be the most effected taxa.

> **COMMENT 12-7**: While the assessment factor derived by the new method makes more sense because it is pesticide specific and does not contain data from other chemical classes, there are several concerns with the new assessment factor methodology.
>
> 1) The pesticides selected for this analysis do not reflect the full set of pesticides for which there is at least some data for. Also, while this procedure uses the method of Host et al, which is the basis for the U.S. EPA Great Lakes Initiative Tier II Assessment Factors, both methods are limiting in that they are not inclusive of smaller data sets. The authors are encouraged to look for potential calculation methodologies that may incorporate smaller data sets. This will ensure a sound scientific basis for the use of this type of factor in that it is not biased by a data richness requirement.

RESPONSE TO COMMENT 12-7: **UCD** The data sets used for this procedure were used because they were reviewed by EPA and already existed in a compiled format. Although, it would be desirable to represent more pesticides, to assemble and evaluate new data sets would be a great deal of work, similar to deriving new criteria. The method has been developed to revise the assessment factor to incorporate additional data as it is developed.

> **COMMENT 12-8**: The additional factor of 10 on top of a factor of 57 (essentially dividing the single taxon acute value by 570), may prove to be problematic in that its application is subjective and not based on objective decision rules or criteria. While the goal is ecosystem protection, there needs to be a balance between criteria so that criteria do not go well beyond there intended use and become an onerous burden on the regulated community. Experience with the GLI Tier II criteria values has demonstrated that these types of values are not widely accepted, since the methodology has been in existence for 10 years, and has seen only limited regulatory use.

RESPONSE TO COMMENT 12-8: **UCD** The application is not subjective. If there is only that one datum from which to derive a criteria then the assessment factor 570 will be applied. These assessment factors were derived from actual pesticide data sets. Appling a factor of 57 to the one required invertebrate value resulted in an under-protective criteria as compared to the full data set for two of the eleven data sets. See section 2-3.2.4, Table 2.7.

It is not intended that this procedure be used widely, as most pesticides should have more than this one data point. However, if a criterion is needed by the Regional Board for a pesticide that has such limited data, this is a means to do so and it is reasonable to be so conservative using just one data point.

> **COMMENT 12-9**: The authors propose a default ACR of 12.4 based on the 80th percentile of a distribution of 8 available pesticide ACRs. Three pesticides (chlordane, dieldrin, and lindane) have been banned for all uses in the United States by U.S. EPA). Since these are not current use pesticides, chronic criteria derivation based on a default ACR including these pesticides is problematic in that it does not represent a current understanding of pesticide mode of action and toxicity, which are somewhat reflected through the acute to chronic ratios calculation. To reflect current science, the authors should consider removing chlordane, dieldrin and lindane and recalculating the default Acute to chronic ratio.

RESPONSE TO COMMENT 12-9: **UCD** Although these are not used in California, there may be more compounds that are similar in that they have chronic effects well below the concentrations that cause acute affects. One possibility could be the endocrine effects of atrazine on frogs. See also response to comment 11-39

> **COMMENT 12-10**: The authors could choose to use a more conservative percentile (e.g. 90th or 95th percentile) to ensure protectiveness of the criteria

RESPONSE TO COMMENT 12-10: **UCD** The procedure for deriving this factor was based on an extensive report by Host et al. (1995) in which they described both empirical and theoretical methods for derivation of factors using data sets for all kinds of chemicals. The 80th percentile was calculated in that report, however the decision to use it was from the Great lakes Initiative. Its use seemed reasonable and taken as a best professional judgment in an area where there was not much guidance.

> **COMMENT 12-11**: Chlorpyrifos toxicity increases at higher temperatures (Chapter 4, pp 4-8). Also, Buchwalter et al., 2003 demonstrated that accumulation of waterborne chlorpyrifos by aquatic animals increases with exposure temperature. This is particularly evident for acute exposures. Elevated metabolic rates (particularly for fish) occur at warmer temperatures increase accumulation minutes to hours after initiation of a waterborne exposure. The report cites work that demonstrates rather remarkable increases in chlorpyrifos toxicity at higher temperatures (15-fold decrease in LC50 for rainbow trout from 7 to 18 degrees Celsius).
>
> Since water is a highly regulated resource in the Central Valley due to agricultural needs and practices, the potential for increased chlorpyrifos toxicity due to elevated temperature in the Sacramento and San Joaquin River systems and elevation is potentially a major issue for salmonids (as stated in the document), particularly if chlorpyrifos presence in streams co-occur with increased temperatures and salmonid presence. There is precedent for adjusting chemical criteria based on water quality parameters such as hardness adjustments for some metals, as well as adjusting temperature criteria to protect valuable aquatic resources such as salmonids. Chlorpyrifos criteria adjusted for temperature is worthy of consideration and should be investigated further.

RESPONSE TO COMMENT 12-11: **UCD** The issue of temperature dependence of chlorpyrifos toxicity was also brought up during peer review and is discussed in the draft chlorpyrifos criteria document in section 4-10.0. While there is evidence of temperature effects on chlorpyrifos toxicity, there are not data for enough species to adequately quantify the relationship at this time. In regards to salmonids, the lowest reported temperature affected toxicity values were above the proposed criteria. Specifically, cladocerans are far more sensitive to chlorpyrifos than salmonids, so any criteria that are protective of cladocerans will also protect salmonids. In addition, cladocerans are generally tested at 25 ºC so that effect levels reflect warmer temperatures. Therefore temperature related criteria adjustments are not are not recommended for the criteria at this time. The question of temperature dependence is a topic that should be revisited as chlorpyrifos criteria are updated.

2.13. Comment Letter 13 – Debra Denton, U.S. EPA

COMMENT 13-1: The exhaustive review of the literature on criteria development as discussed in this report would be extremely valuable to most States in developing criteria.

RESPONSE TO COMMENT 13-1: **UCD/RB** Comment acknowledged.

COMMENT 13-2: One point in particular, I believe will go a long way in deriving either regulatory water quality criteria and developing TMDL numeric targets is the clear and transparent process of reviewing data for its relevance, acceptance and documentation to be considered and reviewed in deriving an individual criteria.

RESPONSE TO COMMENT 13-2: **UCD/RB** Comment acknowledged.

COMMENT 13-3: EPA agrees that when using NOEC data that a consideration of the summary statistic metric, minimum significant difference (MSD) or PMSD should be reported as a measure of within-test variability. I suggest adding the following references to support this (U.S. EPA 2002b; Denton et al., 2003). EPA (2002b) provides a recommendation to implement and evaluate MSD when using hypothesis driven techniques along with power and effect size analyses in Appendix B.

RESPONSE TO COMMENT 13-3: **UCD** This supporting information has been added to section 2-2.1. 2.

COMMENT 13-4: Section 2.1.1.3, Recognizing Section 2.1.1.3 is not an exhaustive list for discussion on nontraditional endpoints, I would suggest additional discussion on biomarker's such as stress proteins which have been linked to abnormal development and larval sturgeon and energetic studies which have demonstrated an increase in energy expenditure to juvenile steelhead trout as discussed in Oros and Werner (2005).

RESPONSE TO COMMENT 13-4: **UCD** The original references (Werner et al. 2007 and Viant et al. 2004) on those two effects were reviewed and a firm link relating the endpoint in question (heat shock protein and metabolic/energetic condition) to a long term detrimental effect was not found, so these have not been added to that section (now section 2-2.1.3). Additionally these studies were on the effects of thermal stress, to which HSP proteins are a direct response. From this information, it is not clear that a pesticide exposure would produce a similar detrimental response.

COMMENT 13-5: I agree for those pesticides with log Kows, between five and seven that feeding routines should be minimized in order to avoid interactions with food particles.

RESPONSE TO COMMENT 13-5: **UCD/RB** Comment acknowledged.

COMMENT 13-6: I agree that the inclusion of EPA's interspecies correlation estimation software be included for this new methodology to address the potential effects with threatened and endangered species.

RESPONSE TO COMMENT 13-6: **UCD/RB** Comment acknowledged.

COMMENT 13-7: Section 2.3.5.2, the discussion on pesticide mixtures are known as well studied, I suggest adding the reference Lydy et al. (2004) which provides an exhaustive literature review of known pesticide to pesticide mixture interactions. I agree with the statement that there is really no way to derive criteria for all the potential mixtures of pesticides that would occur in a waterbody. Therefore, the Central Valley Regional Board 's approach of applying their Basin Plan's additivity formula (toxic unit approach detailed in 3.5.2.1.1) provides a valid approach when dealing with known pesticides located in a given water body. As has been done successfully in the development of the joint diazinon and chlorpyrifos TMDLs for the San Joaquin River, Sacramento and Feather Rivers and Delta.

RESPONSE TO COMMENT 13-7: **UCD** This supporting information has been added to section, now section 2-4.2.

COMMENT 13-8: I agree that having a systematic way of reviewing ecotoxicity data and having a detailed data summary table are paramount to rating the quality of a given study including all the factors listed in this section and provided in Figure 3.3.

RESPONSE TO COMMENT 13-8: **UCD/RB** Comment acknowledged.

COMMENT 13-9: I agree that evaluating ecotoxicity data on the three areas of relevance, documentation and acceptability are crucial to criteria development. This point alone will go a long way in having less contentious criteria development, if it is clear that the data being considered has met these factors and therefore should be included in the database for that individual pesticide criteria document.

RESPONSE TO COMMENT 13-9: **UCD/RB** Comment acknowledged.

COMMENT 13-10: I agree that the minimum taxonomic requirements need to be specified for the minimum data sets.

RESPONSE TO COMMENT 13-10: **UCD/RB** Comment acknowledged.

COMMENT 13-11: I agree with the statement, "Criteria must be protective of aquatic life, and therefore must err on the side of conservatism when data are lacking." This section discussing the application of assessment factors for those data sets that do not meet the minimum data requirements to derive criteria using the SSD approach is most likely necessary, especially in light of the fact that there are fewer pesticide criteria and numerous pesticides which are used and potentially found in water bodies. Therefore, the need exists in the regulatory process in which to establish criteria with fewer data points and is recognized in other countries methodology is cited in Table 2.5.

RESPONSE TO COMMENT 13-11: **UCD/RB** Comment acknowledged.

COMMENT 13-12: Providing Table 3.1 with examples of sources is very helpful for the pesticide review process, and recognizing this is not an

exhaustive list, however, I would consider adding USDA's Chemfinder database in Table 3.1.

RESPONSE TO COMMENT 13-12: **UCD** The Chemfinder database was also added to this table.

COMMENT 13-13: The methodology steps are clearly expressed based on the supporting technical information in Chapter 2.

RESPONSE TO COMMENT 13-13: **UCD/RB** Comment acknowledged.

COMMENT 13-14: The methodology expresses the criteria in magnitude, duration and frequency components consist with existing U.S. EPA 1985 guidelines.

RESPONSE TO COMMENT 13-14: **UCD/RB** Comment acknowledged.

2.14. Comment Letter 14 – Nassar Dean, Western Plant Health Association

COMMENT 14-1: The authors are to be commended for their comprehensive review of the current state of the science regarding the derivation of numeric water quality criteria to protect aquatic life. It is obvious they understand the fundamental questions that need to be answered in this process, and the new proposed method reflects this understanding and presents an interesting synthesis of the best aspects of the existing methodologies according to the professional opinions of the authors

RESPONSE TO COMMENT 14-1: **UCD/RB** Comment acknowledged.

COMMENT 14-2: Why is the methodology restricted to pesticides and not intended for application to all toxic constituents? Each of the existing methods reviewed in Phase I of the project is generally applicable to all chemicals and is not limited to the regulation of pesticides. What are the policy and legal implications of such a limited scope?

RESPONSE TO COMMENT 14-2: **RB** As discussed in the Central Valley Pesticide Basin Plan Amendment Scoping Report, the Regional Board has the authority to regulate those biological stressors that are associated with the

discharge of waste. Pesticides are found in discharges of waste and some of those pesticides are known stressors to the aquatic community, as suggested by the number of pesticide listings on the Clean Water Act Section 303(d) list. Pesticide use information from the Department of Pesticide Regulation suggests that pesticide use is ubiquitous in agricultural and urban settings. Since pesticides are intentionally introduced into the environment to control undesirable plants or animals, it is important to ensure that the discharge of those pesticides does not impact non-target organisms in surface waters. The history of water quality impacts associated with pesticides highlights the need to maintain regulation of pesticide discharges as a high priority.

The focus of this effort on pesticides does not suggest that pesticides are the only stressors to aquatic life beneficial uses. Changes in the natural hydrology of Central Valley streams, along with other stressors, are also likely to impact the aquatic biology. However, an evaluation of all stressors suggests a research program that is beyond the scope of a Basin Planning effort. In addition, the method is not generally applicable to all chemicals. As discussed in response to comment 8-4, there are elements of the methodology that are specific to pesticides and would require additional development to be more generally applicable.

> **COMMENT 14-3**: What is the protection goal? If it is for the protection of all species, then what is the justification for this decision? Protection of 95% of the species (U.S. EPA 1985 method)? The widely accepted concept that aquatic ecosystems can tolerate some stress and therefore protection of all species always and everywhere is not necessary (U.S. EPA 1985 method) is not discussed in the context of protection goals that will be met by the new proposed method.

RESPONSE TO COMMENT 14-3: **UCD** Please see response to comment 2-5.

> **COMMENT 14-4**: The need for a new method is not explained. Existing methods are capable of dealing with both robust and sparse toxicity data sets. Are there deficiencies in the methods used by U.S. EPA and CDFG? If so, these deficiencies should be stated explicitly.

RESPONSE TO COMMENT 14-4: **RB** See responses to comments 1-1 and 1-4.

COMMENT 14-5: It is unclear why the focus appears to be on developing a new national scope criteria derivation process, instead of focusing on how available tools can best be applied or adjusted to take into account the site-specific or regional ecosystem characteristics found in the Central Valley. It would appear that this approach would be more consistent with the regional board's regulatory mandate. More benefit would be gained by expending limited resources on Central Valley specific needs, instead of changing a widely established and well accepted methodology.

RESPONSE TO COMMENT 14-5: **RB** See responses to comments 2-2 and 2-4.

COMMENT 14-6: Specific procedures in the new proposed method exclude data that may have been used previously in existing methods, resulting in greater uncertainty in the final acute and chronic criteria for pesticides. Examples include requiring only 5 toxicity data points in an SSD compared to 8 for the U.S. EPA 1985 method, exclusion of outliers in SSDs for study results that passed earlier data quality evaluation, exclusion of community level data from mesocosm studies, exclusion of species from families found outside North America, and assuming registrant GLP guideline studies are unavailable. In particular, the last point is easily addressed by working with DPR and/or U.S. EPA OPP to obtain toxicity data submitted for registration.

RESPONSE TO COMMENT 14-6: **UCD** The justification for using a minimum of only 5 data points is discussed in response to comment 6-7. The requirement of 5 data does not exclude other data from being used. Outliers are no longer excluded; see response to comment 6-13.

EPA doesn't include mesocosm data in their criteria calculation in the 1985 guidelines, although, similar to the proposed procedure they do consider this data in the final criteria determination. EPA did use plant mesocosm data to derive their freshwater atrazine criteria, but not all data sets are so rich as the atrazine data set. EPA 1985 guidelines exclude species not in North America, so more data may be included in the proposed method because it includes species in all taxonomic families in North America (this is discussed in Ch2, see below).

The methodology has been revised to require requesting DPR and EPA unpublished data, including registrant data as the first data source for deriving new criteria. See response to comment 4-1.

RB Where available, Regional Board will utilize data generated from the registration process. However, as discussed in detail in response to comment 1-1, registration requirements do not reliably provide for sufficient data to perform the US EPA 1985 method. As a result, the Regional Board is interested in a methodology that can utilize the data that is available. Also see response to comment 12-4.

> **COMMENT 14-7**: The use of assessment factors greatly increases the possibility of overestimating risk as reported in the cited Chapman et al. (1998) article. There may be instances where assessment factors are needed due to limited data availability. However, considering every high-quality data point and multiple lines of evidence (lab and field) should minimize the cases where an assessment factor approach is required.

RESPONSE TO COMMENT 14-7: **UCD** The use of assessment factors increases the possibility of overestimating risk and should only be used where there is insufficient data to perform the SSD procedure. Field data will generally not substitute for a lack of the required species specific data.

> **COMMENT 14-8**: Key aspects of the proposed method are characterized by the authors as being required by policy considerations rather than being selected on scientific merit. These include types of data to be considered, one-way adjustment of final values, selection of points on a distribution and associated confidence bounds, determination of assessment factors, and frequency of exceedances. In each of these examples it would be more helpful to identify and characterize different choices supported by data and the state of the science and communicate the uncertainties so that risk managers can make appropriate decisions in relation to clearly communicated protection goals. Such decisions cannot be made independently by scientists dealing with only the risk assessment phase of the risk analysis process (risk assessment, risk management, risk communication).

RESPONSE TO COMMENT 14-8: **UCD**

The choices to include certain procedures in this method were based on a combination of best professional judgment, the science available, and use by a major country (policies). Policy did dictate that a procedure was needed to derive protective criteria with datasets of varying sizes. Other than that, policy was not a major driver in the decision to use the specific procedures for: types of data to be considered, one-way adjustment of final values, determination of assessment factors. Use of the 5th percentile and frequency of exceedances initially came

from a past policy use and best professional judgment, but has been backed up by studies. The factors driving the selection of the individual components of the methods are already well described in Phase I and Ch 2 of Phase II. Also, a new section discussing of the limitations, assumptions and uncertainties in the method has been included (Section 3-8).

RB See response to comment 11-7.

> **COMMENT 14-9**: There are two major factors that must be taken into account when recommending allowable exceedances to meet clear protection goals. These factors are 1) the return frequency of an exceedance, and 2) the magnitude of the exceedance above the criterion value. In the absence of protection goals in the present version of the report, we assume sustainability of aquatic communities associated with the relevant designated uses is the key policy interest of the State. For a method that will only be applied to pesticides there needs to be much more detailed interpretation of the existing data, including numerous microcosm and mesocosm studies, to make recommendations related to the protection goal. Referring primarily to old data and interpretations of point source emissions of industrial and waste chemicals is not necessarily applicable to relatively infrequent pulsed exposures typically observed for pesticide residues in water (non-point source discharges). Likewise, generalizing from recovery times following ecological disasters does not relate to generally small inputs of pesticide mass from diffuse sources.

RESPONSE TO COMMENT 14-9: **UCD** These studies examined mostly one time or multiple (pulsed) pesticide applications and run off from applications, not point source discharges of industrial and waste chemicals. The focus of this review was studies of brief, mild, limited scope excursions, not ecological disasters. Please see responses to comments 14-63, 14-64 and 14-65.

> **COMMENT 14-10**: Some consideration should be given to the ability of an aquatic ecosystem to tolerate slight exceedances, since most species will not be affected (as evidenced by SSDs). Moreover, those species affected at or near the criterion level should be examined for ability to recover from an exceedance in terms of generation time and immigration potential. Lastly, the binomial approach for listing and delisting impaired water bodies used by the California State Water Resources Control Board (SWRCB) was not evaluated. Therefore, making a rigid recommendation to take action in all cases when any level of exceedance occurs above a

highly protective criterion value more than once in a three-year period does not appear to be scientifically justified.

RESPONSE TO COMMENT 14-10: **UCD** Consideration was given to the ability of an aquatic ecosystem to tolerate slight exceedances. If the commenter is calling for species specific frequencies of exceedances, this very likely would require more information than available. For more discussion, please see response to comment 11-9.

RB See Regional Board response to comment 11-10 for information on the binomial approach.

COMMENT 14-11: Although the authors discuss the use of the 5th percentile in other existing methods, there is no rigorous analysis supporting the decision to recommend this point on the distribution for determining criteria in the new method.

RESPONSE TO COMMENT 14-11: **UCD** See response to comment 11-11.

COMMENT 14-12: The authors should describe the scientific rationale for applying a factor of 2 to the 5th percentile value?

RESPONSE TO COMMENT 14-12: **UCD** See response to comment 11-11.

COMMENT 14-13: The authors should recognize that species with more sensitive endpoints than the 5th centile value may or may not be affected at the population level.

RESPONSE TO COMMENT 14-13: **UCD** This is possible, but also hard to predict. The goal is protection of all species, so if there is good evidence that a species may be affected at a level below the 5th percentile (actually the 5th percentile divided by two), the criterion should be lowered so that is becomes protective.

COMMENT 14-14: For pesticides where microcosm and mesocosm studies are available, the 5th centile value can be checked for an adequate or excessive level of protection by comparing it to population and community level responses.

RESPONSE TO COMMENT 14-14: **UCD** See response to comment 4-3.

COMMENT 14-15: When considering the consistency in approach for all compartments (water, sediment, biota, air), the most important aspect probably is sediment quality, since this appears to be the next area where the State will develop regulatory science and science policy. Initiatives underway in the SWRCB related to SQOs should be included, particularly with respect to the need for multiple lines of evidence (MLOE) to reduce uncertainty in determining impairment. Also, dealing appropriately with bioavailability will become even more important for understanding sediment toxicity and impairment.

RESPONSE TO COMMENT 14-15: **RB** See response to comment 11-12.

COMMENT 14-16: The 1-h and 4-d averaging periods are consistent with the U.S. EPA 1985 point-source method that established these general recommendations from a review of limited data for chemicals dissimilar to modern pesticides. It is not clear from the discussion in the present report that they are applicable to California NPS conditions. The typical monitoring programs conducted for pesticide residues in surface water do not appear to be compatible with these averaging periods, and historical monitoring data sets may support more appropriate numbers.

RESPONSE TO COMMENT 14-16: **RB/UCD** See response to comment 11-8.

COMMENT 14-17: Why does this methodology apply only to pesticides? All of the referenced established methods are generally applicable to toxic contaminants. What are the policy implications? If the method is solely targeted at pesticides, then it should more fully benefit the data generation, risk assessments, and overall registration process that occurs on a routine basis at the state and federal level. The proposed methodology demonstrates a lack of understanding of the pesticide registration process and the resources that go into it.

RESPONSE TO COMMENT 14-17: **RB** See responses to comments 1-1, 8-4 and 14-2.

COMMENT 14-18: When will the Phase-III report be available and what is the process associated with it? Will and if so, when will the Basin Plan be amended to incorporate the new methodology? Will and if so, when

will the regional board post responses to these comments on their website?

RESPONSE TO COMMENT 14-18: **RB** See response to comment 11-13.

COMMENT 14-19: On page 2-1, the authors state 11 other pesticide data sets were used from EPA, but only 9 references are given.

RESPONSE TO COMMENT 14-19: **UCD** See response to comment 11-15.

COMMENT 14-20: While it is doubtful that the situation will occur routinely, exclusion of aquatic toxicity test results from species in non-North American families is unnecessary. Taxonomy is an imperfect predictor of relative sensitivity, and under the current proposed scheme each study should be judge on its quality. Judging relevance of specific species as a surrogate for the species found in a specific ecosystem is valid but is generally reserved for site-specific criteria development.

RESPONSE TO COMMENT 14-20: **UCD** Judging the relevance of specific species as a surrogate for the species found in a specific ecosystem needs to be done carefully, so families not native to North America were excluded. It would be difficult to rationalize enforcement of a criteria that is largely influenced by a species from a non-North American family, so they are not included in the criteria derivation calculation for this method. However such an organism would still be included as supplemental information. See also response to comment 11-6.

COMMENT 14-21: The definitions on page 2-2 should include Acute methods for plants

RESPONSE TO COMMENT 14-21: **UCD** Please see response to comment 11-16.

COMMENT 14-22: While the MATC is acceptable, U.S. EPA's recently released Cu criterion document highlighted the EC20, where calculable, as generally corresponding to the MATC and as being the preferred chronic endpoint. The same should be done in the CA methodology.

RESPONSE TO COMMENT 14-22: **UCD** The use of the EC20 in the EPA Cu criteria (U.S. EPA 2007a) follows the preference for chronic effects data from

regression analysis vs. hypothesis tests, which is discussed in the Phase II report, but the proposed method used hypothesis data for the chronic criterion due to the types of data usually available. In the Cu criteria report, there was a large amount of available chronic data that could be examined thoroughly. However, because the EC20 worked well for this one compound, does not warrant that it be the preferred chronic endpoint for all pesticides at this time.

Since most available chronic data are as NOEC/LOECs, that is the primary form used in the method. Additionally, as the copper criteria document and the comment states, the resulting chronic values are similar anyway. However, the proposed method does allow for ECx data to be used "only if studies are available to show what level of x is appropriate to represent a no-effect level."

COMMENT 14-23: It is unclear how the use of non-traditional endpoints may be used to derive criteria if those endpoints have been adequately linked to effects on survival, growth and reproduction or population tests, and the criteria developed from them, are predictive and protective of ecosystem. Will and if so, when will those parameters be developed? Who makes this very critical decision on the use of non-traditional endpoints for criteria derivation? Is it a panel of experts and/or a regional board scientists?

RESPONSE TO COMMENT 14-23: **UCD/RB** See response to comment 11-17.

COMMENT 14-24: Multi-species data from field and semi-field studies should be more fully incorporated in the criteria development process. These are our best available tools to assess whether single species laboratory microcosm and mesocosm data should used in the criteria derivation process. If it is available, they should be used to adjust any criterion developed based on single species laboratory data, as the science dictates (either higher or lower).

RESPONSE TO COMMENT 14-24: **UCD** See response to comment 4-3.

COMMENT 14-25: Inordinate concern is expressed for the role of dietary exposure as an important route of exposure in aquatic tests. Most research has shown that exposure through the water is the dominant route of toxicant entry for most substances and aquatic organism.

RESPONSE TO COMMENT 14-25: **UCD** Diet can be a significant route of exposure, depending on the compound (see Phase I report section 7.1.2). However, that section of the Phase 1 report and the section the commenter is referring to (now numbered as section 2-2.1. 5) does state that water only assessments are best, primarily because there is not a good way to incorporate diet into criteria derivation at this time. This section only makes the suggestions to minimize effects by not feeding in acute tests, which is fairly standard, and that if these criteria are shown to be under protective, to then consider diet as an important route.

COMMENT 14-26: How do the data evaluation criteria affect use of the EPA ECOTOX database?

RESPONSE TO COMMENT 14-26: **UCD** This method should not affect the use of the ECOTOX database. Some of the categories were based on the content of the database. The database is a suggested source for finding studies, but the original studies will be evaluated by the data evaluation process in the proposed method.

COMMENT 14-27: Registrant data are assumed not to be available. However, for all pesticides with outdoor uses, a full data set of physical chemical data is available. This should be the first source of information in this area.

RESPONSE TO COMMENT 14-27: **UCD** The method has been revised to include checking for registrant data with EPA and DPR. See response to comment 4-1.

COMMENT 14-28: Linkage of endpoints to survival, growth, or reproductive effects is appropriate.

RESPONSE TO COMMENT 14-28: **UCD/RB** Comment acknowledged.

COMMENT 14-29: It is understandable that there is a preference to rely on aquatic toxicity data generated using technical grade test material, it is unwise to have a blanket exclusion of tests conducted with formulations. Formulation tests may add additional species to the process, which is very

important. Formulation studies should be considered at least as supplemental information.

RESPONSE TO COMMENT 14-29: **UCD** They are considered as supplemental information. All data that don't rate high enough to be part of criteria calculation are tabulated. These values are checked after the criterion is derived to make sure the criteria will be protective.

COMMENT 14-30: It is agreed that studies that greatly exceed the solubility limit are less than ideal, but they do have some limited utility. For example, if an insecticide has a set of algal studies with values greater than the solubility limit, this data is an indication that algae are unlikely to be sensitive to the material and this factor needs to be included in the criteria derivation process.

RESPONSE TO COMMENT 14-30: **UCD** The method allows consideration of toxicity testing conditions that use concentrations up to 2 times greater than the geometric mean of available water solubility data, but the reported LC50 or LEOC/NOEC is with 2x solubility. When the reported toxicity values are greater than 2 times the solubility, they cannot be used for criteria calculation or in the toxicity supplemental information. These species are apparently insensitive to the toxicant. If an LC50 is derived anyway it may be from another indirect effect, such as the substance forming an oily layer at the water surface that blocks oxygen diffusion. This would not be appropriate to include the toxicity data for that compound.

COMMENT 14-31: While endpoints reported a “<” or “>” values are not useful for regression analysis, they are useful in ranking a species relative sensitivity. Such studies, if otherwise valid, should be included in the total “n” (number of studies) available, either when ranking studies for criteria derivation based on an SSD approach, or using the assessment factor approach.

RESPONSE TO COMMENT 14-31: **UCD** See response to comment 8-10.

COMMENT 14-32: The rationale behind using the 75th percentile of scores for the reliability rating is needed. Chlorpyrifos may not be a good dataset to use for this benchmark, since this is a fairly rich dataset and most other pesticide toxicity data sets will be less extensive.

RESPONSE TO COMMENT 14-32: **UCD** See response to comment 5-5.

> **COMMENT 14-33**: It is highly unlikely that any pesticide will be registered in the US (or most other countries) based solely on two data points. The statement on page 2-15 is misleading and illustrates a lack of understanding of the intensive registration process that pesticides undergo. No other group of materials has as much ecotoxicology data generated prior to commercial use, as pesticides do.

RESPONSE TO COMMENT 14-33: **UCD** While overall there are many data required for pesticide regulation, there are really only three required studies on acute toxicity to freshwater aquatic life [emphasis added]. See below for details, the underlined requirements are which are useable for this purpose.

From 40 CFR Part 158.630: "In general, for all outdoor end-uses, including turf, the following studies are required: Two avian oral LD_{50}, two avian dietary LC_{50}, two avian reproduction studies, two freshwater fish LC_{50}, one freshwater invertebrate EC_{50}, one honeybee acute contact LD_{50}, one freshwater fish early-life stage, one freshwater invertebrate life cycle, and three estuarine acute LC_{50}/EC_{50} studies -- fish, mollusk and invertebrate. All other outdoor residential uses, i.e., gardens and ornamental will not usually require the freshwater fish early-life stage, the freshwater invertebrate life-cycle, and the acute estuarine tests."

> **COMMENT 14-34**: Exclusion of some taxa from the EPA list is justified by the authors by citing insensitivity to pesticides. As all taxa would need to represented in the community its exclusion would seem to bias the statistics. Consideration should been given to molluscicides. If not, then what is the rationale for it's exclusion? WQC are supposed to represent the entire community, and excluding or including species based on perceived sensitivity is scientifically unjustifiable.

RESPONSE TO COMMENT 14-34: **UCD** Data on these taxa are in no case excluded from the SSD. They are, however, not required for criteria derivation. Which taxa are included depends on data availability.

The target taxa are not necessarily the most sensitive species. Mollusks are usually much more insensitive to pesticides than other species, sometimes even to chemicals used to exterminate mollusks, and there are a lot of other species that are not required. Not every taxa can be required for water quality criteria, as

many of these studies would be rare. The absence of mollusk data should not prohibit criteria derivation. Of course all relevant species will be included where data is available. Again, they are just not required.

Mollusks are not specifically required by EPA's 1985 Guidelines either. They are one of several options in the seventh requirement: "a family in a phylum other than Arthropoda or Chordata (e.g. Rotifera, Annelida, Mollusca, etc.)"

> **COMMENT 14-35**: Most standard testing methodologies for benthic invertebrates include sediment. Inclusion of a water column benthic crustacean test requirement is inconsistent with current standard test methodologies to assess the toxicity to benthic organisms.

RESPONSE TO COMMENT 14-35: **UCD** While it is common to also do sediment in benthic invertebrate toxicity tests, this should not be limiting for the proposed method. The use of a benthic crustacean is still a requirement in EPA water quality guidelines and the EPA Office of Prevention, Pesticides and Toxic Substances website has Ecological Effects Test Guidelines for Gammarid (850.1020) and Penaeid (850.1020), two benthic crustaceans, and these are aqueous test protocols without sediment.

> **COMMENT 14-36**: Reducing data to the species rather than genus level is appropriate.

RESPONSE TO COMMENT 14-36: **UCD/RB** UC Davis and Regional Board staff appreciate the commenter support for this element of the method.

> **COMMENT 14-37**: It is understood why the more sensitive life stage is being chosen, but it should be the more relevant life stage for the ecosystem of concern. On this point, the regional board is better served by considering site-specific approaches rather than developing new national scope criteria.

RESPONSE TO COMMENT 14-37: **UCD** Since these criteria apply to whole watersheds, it is unlikely that the more sensitive life stage of a given resident species would not occur in this system. It is difficult to find an example in which this would be a problem. Even with anadromous fishes that live both in the relevant watersheds and outside them in the ocean, such as salmon and stripped bass, the larvae live in fresh water.

RB The Regional Board has neither the authority nor the intent to establish a national criteria derivation process. See Response to comment 2-2.

> **COMMENT 14-38**: Care needs to be taken on when and how "outliers" are removed from the process. Exclusion of data is easily perceived as arbitrary, and a reason (i.e., related to study design or reporting) beyond simply not fitting a specific distribution should be required.

RESPONSE TO COMMENT 14-38: **UCD** Outliers are no longer removed by a statistical test. See response to comment 6-13.

> **COMMENT 14-39**: The protection goal is given as protection of all species in an ecosystem. If and how does this relate directly to the language of the regional board Basin Plan? Please provide WPHA the statutory definition of the term "protection" as utilized by the CVWQCB in regulatory actions.

RESPONSE TO COMMENT 14-39: **RB** See response to comment 2-5.

> **COMMENT 14-40**: Are the pesticide data sets reported in Table 2.1 representative of Central Valley conditions? If not, then what is the rationale and justification for its applicability? Also, the data sets appear to be incomplete. For example, many more species than 9 have been tested with atrazine.

RESPONSE TO COMMENT 14-40: **UCD** These data were taken straight out of EPA criteria documents for the specific chemicals. They are inclusive of, but not restricted to Central Valley conditions. Restricting the data set to only data reflecting Central Valley conditions is unnecessary and would greatly reduce the available data.

The EPA criteria data were used because they were readily available. In Chapter 2 section 2-3.1.1 it states that for the atrazine data set, > or < values were not included because they are only useable in the log-triangular distribution. Without those values there were 9 species mean acute values in table 1 of the atrazine EPA criteria document and those are the values included in the report.

COMMENT 14-41: The authors note that it is "important to minimize violations of distributional assumptions," yet they bias the data set by placing an emphasis on sensitive species, instead of striving for a robust and representative data set. As mentioned above, "< and >" values can be included in the ranking process to determine the total "n", but not used for the regression. This will lead to a more realistic ranking of relative sensitivities. It should also be noted that the importance of distributional assumptions in this application of SSDs (deriving a single point estimate) is debatable, since non-distributional approaches have been shown to work as well. The practicality of the U.S. EPA 1985 method in using the four points close to the 5th percentile should not be dismissed.

RESPONSE TO COMMENT 14-41: **UCD** See response to comment 8-10.

COMMENT 14-42: The comparison of how the data sets fit the various distributions that is summarized in Table 2.3 is misleading. Burr type III distribution represents a family of distributions, so of course it appears to fit more compounds than when compared to individual distributions. Also, while the U.S. EPA standard WQC method cites a log-triangular distribution, actually what is done for criteria development is to use the 4 points close to the 5th centile to derive the criterion. The EPA standard method recognizes a critical point; if we are interested in deriving a point estimate from an SSD, then it is relatively unimportant if overall fit to the distribution is good – the key question is whether the fit is good at the lower end – where the 5th centile is being estimated. There is much practical value in the EPA method of relying on the data points closest to the centile of concern. The confidence interval around the 5th percentile is a better predictor of fit/value of the method when deriving WQC. In cases where the full distribution will be used (such as generation of joint probability curves) then the model fit is of larger importance.

RESPONSE TO COMMENT 14-42: **UCD** The report does discuss how the Burr type III distribution represents a family of distributions, and how, because of that, it has the potential to fit data better. Table 2.3 is used to exemplify this point. This is the last part of section 2-3.1.1, which clearly talks about how it is a family of distributions and approximates some of the other distributions:

'The Burr family of distributions is used in the new methodology for derivation of criteria by the SSD technique because it provides a better fit than the log-triangular distribution in all cases tested, and provides an equivalent or far better fit than the lognormal distribution in most cases. This is expected because the Burr III family of distributions approximates the log-normal and log-triangular distributions (CSIRO 2001)'

The confidence interval around the 5th percentile is a better predictor of fit/value of the method. There is no confidence interval or a measure of fit for the log triangular method, while the proposed method can calculate the confidence interval at the 5th percentile.

See response to comment 8-7 on focusing on using Burr III with all data vs. log-triangular and the sensitive end.

> **COMMENT 14-43**: Rather than prescribe the distribution to use for the pesticide toxicity data, Burr III distribution, why not use the distribution that best fits the data?

RESPONSE TO COMMENT 14-43: **UCD** See response to comment 9-5.

> **COMMENT 14-44**: The 5th centile is characterized as having been validated by field studies, but the details supporting this characterization specifically for pesticides are not provided – were effects found at the 5th centile, or was it shown that the 5th centile from laboratory data is highly protective? For example, it was demonstrated for pyrethroids that the 5th centile was quite protective of aquatic ecosystems: LOEC concentration from multiple mesocosm and field studies with cypermethrin and esfenvalerate corresponded to around the 50th centile of the acute SSD for arthropods, the most sensitive group of species for the pyrethroid. Ecosystems are typically quite robust, especially in comparison to laboratory - based single species data.

RESPONSE TO COMMENT 14-44: **UCD** Section 2-3.1.2 of the Phase II report describes agreement between the 5th percentile and the NOEC, so the 5th percentile approximately equated to the NOEC. There was no report of a large margin of safety.

The reference used to support the comment (Giddings et al. 2001) follows the theory that the effects on an individual population are not significant if the ecosystem function is preserved by another species. However, the protection goal of the method is at the species level so the use of the 5th percentile will be used to protect all species.

For example, in Giddings et al., amphipods were said to decline without recovery at the LOEC for cypermethrin. For the low exposure of esfenvalerate, copeopods were reduced did not recover, chironomids were reduced at all treatment levels, hyalella azteca were reduced at all levels. Since another species was able to replace the ones that were lost, these effects were not viewed as significant by Giddings et al. but these are significant effects by the purpose of the proposed method- to protect all species.

See also response to comment 4-3.

> **COMMENT 14-45**: The authors note Solomon's comment that "…any percentile may be chosen as long as it can be validate against knowledge of ecosystem structure and function," yet it does not appear that this logic was applied to this methodology. Since the Central Valley extends over a limited set of ecosystems, these ecosystem characteristics should be considered when drafting, and applying WQC.

RESPONSE TO COMMENT 14-45: **UCD** There are portions of the method that include adjustments for temperature, pH, and water quality factors. Validating the choice of the percentile for a particular ecosystem would be a very laborious and unfeasible task.

> **COMMENT 14-46**: What evidence will be considered to adjust criteria down from the 5th centile level? Can it be adjusted up if there is supporting evidence? If not, why is the adjustment allowed to be only in one direction? Similarly, what evidence is necessary to move from a median 5th centile estimate to a lower 95% confidence limit estimate?

RESPONSE TO COMMENT 14-46: **UCD** See response to comment 4-3.

> **COMMENT 14-47**: Aggregation of taxa by habitat is mentioned only in passing. A more detailed treatment with respect to the U.S. EPA provision to establish site-specific criteria would be helpful. The more that a WQC can include site-specific factors, such as species or important water quality parameters, the more relevant the WQC will be for the ecosystem of concern. Condemnation of the U.S. EPA's standard WQC derivation methodology as the "most often criticized" is a bit unfair. Since it is the first widely applied distributional (before the term "SSD" was common) based criteria development methodology, it is not surprising that it is the one most often criticized/cited.

RESPONSE TO COMMENT 14-47: **UCD** Aggregation of taxa by habitat is not given much consideration, because the lack of available data precludes this approach. Water quality parameters are addressed in the method in sections 3-5.0 through 3-5.3. The section of the report that the commenter describes as unfair has been revised.

> **COMMENT 14-48**: The authors should provide references for assumptions that "apply to all SSD models," since it is not clear that they are indeed universal. In listing assumptions common to all SSD models, there is mention that protecting the most sensitive species will protect all species in an ecosystem. This is very conservative and therefore is best suited as an indication of potential impairment of the biological community. Basic ecology suggests that most ecosystems are relatively robust and can tolerate changes in species composition without changes in ecosystem function. Nowhere is there discussion of the relationship between protection and impairment. Biological data from specific water bodies are necessary to confirm impairment. Are biological criteria under consideration? If not, what is the rationale for its exclusion?

RESPONSE TO COMMENT 14-48: **UCD** The goal of the proposed method includes all species in an ecosystem (see response to comment 2-5). Protecting ecosystem function is not the same. The header for list of SSD assumption was revised to indicate that the assumptions refer to the 3 methods being considered in this section (Section 2-3.1.5.1).
The goal of this method is not to confirm impairment, but to develop criteria that are protective of aquatic life. Biological criteria are beyond the scope of this methodology, and are unnecessary to establish protective criteria. See also response to comment 4-3.

RB The scope of work was developed to generate criteria that would be protective of aquatic organisms, consistent with existing Board policy (see response to comment 2-5). The scope of work does not include consideration of the impairment status of a waterbody. Impairment is determined according to the State Board's listing policy (SWRCB 2004), which includes consideration of biological criteria.

> **COMMENT 14-49**: Assumptions specific to the U.S. EPA method include "aquatic ecosystems can tolerate some stress and occasional adverse effects, therefore protection of all species at all times and places is not

necessary." The assumptions listed for the RIVM and ANZECC methods do not address this point. The proposed method also does not explicitly discuss this point. What is the author's perspective and rationale?

RESPONSE TO COMMENT 14-49: **UCD** The scope of work is to prevent detrimental physiological responses in aquatic organisms, not simply protection of aquatic ecosystems. However, the method does recognize the ability of ecosystems to tolerate some stress and accounts for this by establishing allowable exceedance rate instead of absolute maximum levels. The protection goal has been clarified; see response to comment 2-5.

COMMENT 14-50: In the section on comparing the lowest value in each data set with the resulting criterion to determine whether or not criteria are protective, there is no discussion of the uncertainty associated with testing of different species for each chemical. Presumably this is never done in a manner that systematically attempts to identify the most sensitive taxa, what is the significance of this comparison with actual protection levels? Comparing these values to field and mesocosm studies would probably be a better measure of whether they are protective.

RESPONSE TO COMMENT 14-50: **UCD** See response to comment 4-3. Section 3-8: 'Assumptions and limitations of the method' has been added, which lists some of the sources of uncertainty associated with toxicity testing and criteria extrapolation.

COMMENT 14-51: Why has dividing the 5th centile by two been kept in the new proposed methodology? This appears to be a holdover from a method that the authors have criticized elsewhere. Do other regulatory authorities divide the 5th centile by 2? If so, please name them and their relevance to this issue.

RESPONSE TO COMMENT 14-51: **UCD** See response to comment 11-11. The safety factor is only used for acute data. Most other agencies use primarily chronic/ no effect data and do not derive a separate acute criterion, so no safety factor is used.

COMMENT 14-52: Instead of developing criteria for pesticides using a new methodology, either a SSD or assessment factors approach, it is more appropriate to rely on the recently released aquatic life benchmarks by U.S. EPA OPP for pesticides to meet the needs of the regional board.

These are values already applied in the regulatory context and use all the data available to the U.S. EPA OPP registration process.

RESPONSE TO COMMENT 14-52: **RB** See response to comment 1-5.

COMMENT 14-53: The authors state, "Each of the points raised by Chapman et al. (1998) need to be evaluated in the context of water quality criteria derivation, which is not the same as ecological risk assessment." Such a distinction is not consistent with recent examples of pesticide criteria derivation such as the U.S. EPA OPP and OW cooperation on the draft ambient aquatic life water quality criteria for atrazine. These criteria draw on numerous ecological risk assessment publications and incorporate risk assessment principles into the final expression of criteria conditions.

RESPONSE TO COMMENT 14-53: **UCD** It is not the authors' intention to say that risk assessment principles should never be applied to water quality criteria. Indeed, many concepts and procedures are the same; however, there are some differences as discussed in the report in Section 2-3.2.1. Briefly, ecological risk assessment seeks to estimate risk based on a specific set of exposure and effects data, usually for a specific site. A numeric water quality criterion, on the other hand, is one number that aims to be protective of aquatic life for the range of sites, and therefore must err on the side of conservatism when data are lacking.

COMMENT 14-54: Since pesticides undergo a risk assessment in the licensing process under FIFRA, and the proposed methodology is currently intended to apply only to pesticides, there is a need to better harmonize the two systems of evaluation to better serve the entire regulatory process at the federal and state levels.

RESPONSE TO COMMENT 14-54: **RB** The Regional Board will continue to look for opportunities to coordinate our efforts with other agencies. However, harmonization of regulatory programs with other agencies is beyond the scope of this project. Also see response to comment 2-1 and 5-11.

COMMENT 14-55: Note, for example, the recent release of aquatic life benchmarks by OPP, which are characterized as "only indicators." This characterization would appear to require consideration of risk assessment principles. If the regional board does not wish to do this at the stage of criteria derivation, then it is clearly necessary to bring in other lines of

evidence to reach impairment decisions - which should then be based on risk characterization procedures.

RESPONSE TO COMMENT 14-55: **RB** Determination of impairment status of surface waterbodies is being performed as part of the 2008 303(d) List update and is beyond the scope of this project.

COMMENT 14-56: The authors state "all criteria are extrapolated values." Therefore they are subject to uncertainty, and additional lines of evidence are needed to determine whether specific local aquatic communities are actually impaired when any numeric criteria are exceeded. This is particularly true when the numeric criteria are quite low, the allowable exceedance frequency is set at three years in all cases, and no consideration is given to the magnitude of the exceedances.

RESPONSE TO COMMENT 14-56: **UCD** See response to comment 4-3.

COMMENT 14-57: The use of toxicity data from the daphnid family for limited toxicity data sets may be overprotective for some chemical classes and under-protective for others, depending on taxa that are sensitive to a particular mode of action.

RESPONSE TO COMMENT 14-57: **UCD** See response to comment 11-24.

COMMENT 14-58: Table 2.6 provides a clear example of why limited toxicity data (n<5) should never be used to establish criteria. The assessment factors are derived from insecticide data. Are these generally applicable to all classes of pesticides? If so, what is the rationale for its use?

RESPONSE TO COMMENT 14-58: **UCD** See response to comment 6-15 on use of less than 5 toxicity data.

The default acute to chronic ratio (ACR) and assessment factor (AF) methodology were formulated with data from organic insecticides. Some molluscicides, miticides, fungicides have similar properties as well and these factors would serves as a reasonable means of estimating criteria in theses cases. The method also intends for the default ACR and the AFs to be living values that will be recalculated when new ACRs or pesticide data sets are available (section 3-3.3 and 3-4.2.3).

One class of pesticides for which an alternative to the AF and default ACR procedures is included is herbicides (or if plants are most sensitive). A different procedure is included for herbicides since no plant data were included in the AF calculation and an ACR cannot be used with plant data because there are generally no acute plant values (see response to comment 11-16 for details). Where other pesticides exhibit similar data limitations some professional judgment may be needed.

RB See also response to comments 1-1 and 6-15.

> **COMMENT 14-59**: Marine organisms should only be used if salinity does not affect the toxicity of the pesticide. For example, salinity can affect the toxicity of metals used as active ingredients.

RESPONSE TO COMMENT 14-59: **UCD** Using data for marine species is only permissible for the derivation of an ACR.

> **COMMENT 14-60**: The Great Lakes guidance document select the 80th percentile as a default value of ACRs. It should be stated clearly that if an ACR is available, for example chlorpyrifos, then this ACR is used and not the default value of 12.4. The very large ACR for lindane is suspect. The ACRs are derived from insecticide data. Are these generally applicable to all classes of pesticides? If so, then why?

RESPONSE TO COMMENT 14-60: **UCD** The method does call for use of pesticide specific data if adequately available, before using the default ACR. See response to comment 14-58 on the use of the default ACR for different classes of pesticides. See response to comment 11-27 on the ACR for Lindane.

> **COMMENT 14-61**: The reference supporting the statement ". . . the chronic averaging period of 4 days has been shown to be long enough to observe the equivalent of chronic toxicity (U.S. EPA 2002c) . . ." appears to be incorrect. It is therefore not possible to evaluate the supporting evidence. Also, this statement may not be true for growth endpoints in longer-lived species such as fish.

RESPONSE TO COMMENT 14-61: **UCD** The reference is correct. It describes how short-term tests of 4-7 days (for algae, daphnia magna, and fathead

minnow) can approximate chronic toxicity. The text has been revised for clarification.

> **COMMENT 14-62**: The comments that chlorpyrifos and diazinon are not fast acting toxicants is not supported by the newly derived chlorpyrifos acute and chronic values which are nearly identical and a previously published EPA diazinon criterion of 100 ng/L for both the acute and chronic criteria .

RESPONSE TO COMMENT 14-62: **UCD** Please see response to comment 11-29

> **COMMENT 14-63**: In the discussion of evidence for ecosystem recovery times on page 2-58, the authors appear not to distinguish between ecological disasters and studies where more environmentally relevant concentrations were investigated. This is an example of omitting consideration of the magnitude of exceedance above a specified level and merely assuming all exceedances will have the same level of impact. This is clearly not the case as evidenced by SSDs.

RESPONSE TO COMMENT 14-63: **UCD** Disasters were certainly not the focus of this section. The author made the point several times that excursions of water quality criteria were not in the realm of ecological disasters. In section 2-3.4 it describes how 3 years would not be sufficient recovery time for an ecological disaster.

> **COMMENT 14-64**: The summary of pulse exposure studies likewise fails to consider dose as a factor influencing time to recovery.

RESPONSE TO COMMENT 14-64: **UCD** This section focused on events resembling mild exceedances. Predicting the outcome of an exceedance varies on many factors as the commenter suggested in this comment and the next one (comment 14-65). These factors and many more were discussed in Chapter 2, section (2-3.4.1).

With all these factors to consider, it would be incredibly difficult and laborious to determine allowable exceedances based on particular doses and durations. Also there is a limited amount of information available to assess such specific questions. The more general question of what is the maximum time needed to

recover for any probable exceedance is much more answerable with the information available.

> **COMMENT 14-65**: The conclusion that a 3-year recovery time is necessary for all excursions above either acute or chronic water quality criteria is not supported by the evidence cited by the authors. Additional interpretations of microcosm and mesocosm studies have been omitted (for example, see Giddings reference for comment 14-44.

RESPONSE TO COMMENT 14-65: **UCD** It is rather difficult to specify a different recovery time for each excursion as discussed in response to comment 14-64.

While the Giddings study does have some good information, it's also lacking recovery for the most sensitive species. This is inconsistent with the method goal (see response to comment 2-5).

In the Giddings study it is stated that most species recover in a few weeks. However, in some cases there was no recovery. For exposure to cypermethrin, ephemenoptera, amphipods and isopods were said to decline without recovery. For esfenvalerate, "copeopods…. did not recover, chironomids…reduced at all treatment levels and showed little or no recovery, tube building shredder… reductions in sensitive taxa were still evident five months after the last esfenvalerate application, hyalella azteca… eliminated….no recovery occurred within the 53-d study."

This study is not one of the best to include because some of the exposures lack a time to full recovery of all species, and the goal of the method is to protect all species.

> **COMMENT 14-66**: Solid Phase Micro Extraction (SPME) technology is probably superior to SPMD devices to characterize bioavailability.

RESPONSE TO COMMENT 14-66: **UCD** Since both of these methods are relatively new, it is difficult to form an opinion one way or another at this time.

> **COMMENT 14-67**: It is appropriate to allow correction for bioavailability when data are available. The precedent clearly has been set to adjust criteria based on water quality factors that have been shown to modify

toxicity. Equilibrium partitioning theory is well establish and should be used where appropriate to modify criteria. To not modify the criteria, based on bioavailability considerations, leaves one with criterion that will have little relevance to the real world.

RESPONSE TO COMMENT 14-67: **UCD** There is a bioavailability section (3-5.1) at the beginning of the water quality effects section that uses an equilibrium partitioning model calculation.

COMMENT 14-68: Additivity should take into account thresholds and use valid measures of effect, such as a relative toxicity approach based on common testing (cited Felsot, 2005 reference) and not water quality criteria based on independent testing.

RESPONSE TO COMMENT 14-68: **RB** See response to comment 11-31.

COMMENT 14-69: The new proposed mixture methods will likely propagate error inherent in the individual components, rendering the methods unsuitable for regulatory decision-making. Also, the expressions are overly complex. The regional board would be better served by investing in biological monitoring to determine the status of aquatic communities if mixtures truly are a concern.

RESPONSE TO COMMENT 14-69: **UCD** These equations are really not very complex. Using them will be much easier than implementing monitoring. The uncertainty that is propagated will not over come the use of these methods as a simple way to provide protection against and synergistic or additive toxicity.

RB Current Basin Plan Policy requires consideration of cumulative effects. Regional Board staff must consider additive effects where the data indicates that it exists.

COMMENT 14-70: Adjusting derived criteria down to protect a most sensitive species results in the criteria resembling even more closely a screening value that will require additional lines of evidence to support decision-making on actual impairment.

RESPONSE TO COMMENT 14-70: **UCD** See response to comment 4-3 and 14-48.

RB Determination of impairment status is beyond the project scope of work. See also response to comment 14-48.

> **COMMENT 14-71**: Does the regional board have the authority to regulate water quality to protect terrestrial wildlife? If so, please describe the statutory and regulatory authority and basis for this process concerning this specific issue.

RESPONSE TO COMMENT 14-71: **RB** Porter Cologne Section 13241 requires that the Regional Board "establish such water quality objectives in water quality control plans as in its judgment will ensure the reasonable protection of beneficial uses." Beneficial Uses, as defined in Porter Cologne Section 13050, "include, but are not limited to … preservation and enhancement of wildlife." The wildlife beneficial use definition is elaborated in the Basin plan (Section II) as "Uses of Water that support terrestrial or wetland ecosystems including but not limited to, preservation and enhancement of terrestrial habitats or wetlands, vegetation, wildlife (e.g., mammals, birds, reptiles, amphibians, invertebrates), or wildlife water and food sources." Porter Cologne Section 13242 requires a program of implementation to achieve the water quality objectives.

> **COMMENT 14-72**: Registrant BCF studies are the most robust data sources. The regional board should work with DPR and U.S. EPA OPP to resolve any concerns they have in this area, since it is addressed in the pesticide registration process.

RESPONSE TO COMMENT 14-72: **UCD** See response to comment 4-1.

> **COMMENT 14-73**: Why is it not possible to adjust derived criteria upward after they are evaluated against field or semi-field data?

RESPONSE TO COMMENT 14-73: **UCD** See response to comment 4-3.

> **COMMENT 14-74**: Does the regional board have the authority to regulate water quality to protect endangered and threatened species? If so, please describe the statutory and regulatory authority and basis for this process concerning this specific issue.

RESPONSE TO COMMENT 14-74: **RB** Yes, Porter Cologne Section 13241 requires that the Regional Board "establish such water quality objectives in water quality control plans as in its judgment will ensure the reasonable protection of beneficial uses." Section II of the Basin Plan includes a beneficial use of Rare, Threatened, or Endangered Species, which is defined as, "uses of water that support aquatic habitats necessary, at least in part, for the survival and successful maintenance of plant or animal species established under state or federal law as rare, threatened or endangered." The Regional Board has not identified specific water bodies that include the Rare, Threatened, or Endangered Use. However, we believe that these species are protected by the water quality objectives that are adopted to protect all the generic aquatic life use categories and the wildlife use. The water quality objectives are set at levels to protect all the species present, including rare, threatened or endangered species.

> **COMMENT 14-75**: The suggestion to use QSARs to protect endangered and threatened species is interesting. The U.S. EPA OW holds national consultations with the Services to evaluate the protection level of existing federal criteria. It would behoove the CVWQCB to track closely this consultation.

RESPONSE TO COMMENT 14-75: **RB** The Regional Board appreciates this suggestion and will consider tracking this process.

> **COMMENT 14-76**: How does the toxicity data screening process developed by the authors compare with the process used by U.S. EPA OW for their development of water quality criteria? To the criteria used for inclusion in the U.S. EPA ECOTOX database?

RESPONSE TO COMMENT 14-76: **UCD** See response to comment 4-2 for comparisons to other criteria. A direct comparison to the screening used by the ECOTOX database has not been done, but the EXCOTOX screening parameters were part of the basis for the screening/ rating system in the proposed methodology.

> **COMMENT 14-77**: The acceptance criteria score of at least 70 based on a maximum score of 100 as described in Table 3.6 is problematic and will allow invalid data to be used for criteria development. For example, a study conducted with an impure chemical used for testing (minus 15) could still obtain a passing score of 85. Another example would result in

an acceptable score of 92.5 with controls that did not meet the acceptability requirements of the method.

RESPONSE TO COMMENT 14-77: **UCD** Studies with relevance score of 70-90 are only to be used as supplemental information, so scores of 85 would not be acceptable for criteria derivation by this method.

The rating system does not allow use of a study completely void of controls. The rating system is set up so that if a study did not describe a control, but reported a control response (or the reverse) it would be acceptable, for the first rating anyway. The second scoring system, also gives points based on use/ description of control, the loss of those points and those form other areas could exclude the study.

COMMENT 14-78: The authors explain the differences in their new lower acute and chronic criteria for chlorpyrifos compared with the EPA or CDFG values by stating that different data sets were used for final calculations. A more detailed analysis should be included that explains how the three specified methods produce differing chlorpyrifos criteria.

RESPONSE TO COMMENT 14-78: **UCD** See response to comment 4-2.

2.15. Comment Letter 15 – Renee Pinel, Western Plant Health Association

COMMENT 15-1: We are concerned that this new methodology and its anticipated use has the potential to effectively insert this Regional Board into the establishment of pesticide use criteria and restrictions in a manner that effectively bypasses and potentially duplicates the existing registration, labeling and federal water quality regulatory structure for these products.

That existing structure carefully integrates the primary registration/restriction/labeling role of the Department of Pesticide Regulation DPR) with roles and enabling methodologies of U.S. EPA and the California Department of Fish & Game. The stated goal of the Regional Board project is to "develop a methodology for derivation of pesticide water quality criteria for the protection of aquatic life in the Sacramento and San Joaquin River Basins." This seems to so directly duplicate or overlap responsibilities within the existing regulatory structure that it raises a number of issues that have yet to be explored.

RESPONSE TO COMMENT 15-1: RB See response to comment 2-1.

COMMENT 15-2: The California Regional Board is taking on the formidable task of developing what appears to be a new national criteria derivation process.

RESPONSE TO COMMENT 15-2: RB See response to comment 2-2.

COMMENT 15-3: The process has not identified deficiencies in the methods used by U.S. EPA and the California Department of Fish and Game.

RESPONSE TO COMMENT 15-3: RB See response to comment 1-1.

COMMENT 15-4: We believe the role of the Regional Board should instead focus on how available tools can best be applied or adjusted to take into account the site-specific or regional ecosystem characteristics found in the Central Valley.

RESPONSE TO COMMENT 15-4: RB See response to comment 2-4.

COMMENT 15-5: The Regional Board is considering methodology focused on pesticides which are well beyond past actions. Regulatory authority for pesticides under California law resides with DPR and we have not seen any documentation that demonstrates a need for the Regional or State Board to assume regulatory oversight of specific constituents such as pesticides within water systems. Other agencies like the Air Resources Board have successfully developed effective working relationships with DPR without developing duplicative and conflicting regulatory processes.

RESPONSE TO COMMENT 15-5: RB See response to comment 2-1.

COMMENT 15-6: We question the regulatory authority of the Regional Board to develop regulatory standards for specific constituents, outside the established pesticide regulatory structure of California.

RESPONSE TO COMMENT 15-6: RB See response to comment 2-1.

COMMENT 15-7: Conflicting standards will open California agencies up to additional challenges, and add confusion and cost to growers who are trying to comply with potentially inconsistent regulations adopted by multiple agencies.

RESPONSE TO COMMENT 15-7: RB See response to comment 1-6 and 2-1.

COMMENT 15-8: To the best of our understanding, at no time during this process has the need for a new methodology been documented.

RESPONSE TO COMMENT 15-8: RB See response to comment 1-1.

COMMENT 15-9: More importantly, the application of the proposed methodology by the Regional Board and its implications for existing pesticide registration and evaluation processes has not yet been deliberated. The likelihood of conflicting regulatory outcomes is significant given the multiple agencies involved. WPHA believes these issues must be thoroughly explored and reconciled among those involved agencies before any new process can be adopted. We are concerned that this evaluation is not happening and urge your attention to and engagement with this matter.

RESPONSE TO COMMENT 15-9: RB See response to comment 1-6.

3.0 References

Aldenberg T, Jaworska JS. 2000. Uncertainty if the Hazardous Concentration and Fraction Affected for Normal Species Sensitivity Distribution. Ecotoxicology and Environmental Safety 46:1-18.

Anderson BS, Phillips BM, Hunt JW, Connor V, Richard N, Tjeerdema RS. 2006. Identifying primary stressors impacting macroinvertebrates in the Salinas River (California, USA): Relative effects of pesticides and suspended particles. Environ Poll 141: 402-408.

Anderson TD, Lydy MJ. 2002. Increased toxicity to invertebrates associated with a mixture of atrazine and organophosphate insecticides. Environ Toxicol Chem 21: 1507-1514.

ANZECC, ARMCANZ. 2000. Australian and New Zealand guidelines for fresh and marine water quality. Report Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australia.

Beaulaurier, D., G. Davis, J. Karkoski, M. McCarthy, D. McClure, M. Menconi. 2005. Amendments to the Water Quality Control Plan for the Sacramento River and San Joaquin River Basins for the Control of Diazinon and Chlorpyrifos Runoff into the Lower San Joaquin River, Final Staff Report. California Regional Water Quality Control Board, Central Valley Region. Sacramento, CA.

CDFG. 1990. Hazard Assessment of the Rice Herbicides Molinate and Thiobencarb to Aquatic Organisms in the Sacramento River System. California Department of Fish and Game (CDFG) Environmental Services Division. Sacramento, CA. Administrative Report 90-1. 1990.

CDFG. 1998. Hazard Assessment of the Insecticide Malathion to Aquatic Life in the Sacramento-San Joaquin River System. California Department of Fish and Game (CDFG) Office of Spill Prevention and Response. Sacramento, CA. Administrative Report 98-2. 1998

CDFG. 1999. Hazard Assessment of the Fungicides Benomyl, Captan, Chlorothalonil, Maneb, and Ziram to Aquatic Organisms. California Department of Fish and Game (CDFG) Office of Spill Prevention and Response. Sacramento, CA. Administrative Report 99-1. 1990.

CDFG. 2000. Hazard Assessment of the Synthetic Pyrethroid Insecticides Bifenthrin, Cypermethrin, Esfenvalerate, and Permethrin to Aquatic Organisms in the Sacramento-San Joaquin River System. California Department of Fish and Game (CDFG) Office of Spill Prevention and Response. Sacramento, CA. Administrative Report 00-6. 2000.

Chapman PM, Fairbrother A, Brown D. 1998. A critical evaluation of safety (uncertainty) factors for ecological risk assessment. *Environ Toxicol Chem* 17:99-108.

CSIRO. 2001. BurrliOZ v. 1.0.13: Commonwealth Scientific and Industrial Research Organization, Austraila.

Deneer, J.W., T.L. Sinnige, W. Seinen and J.L.M. Hermens. 1988. The Joint Acute Toxicity to Daphnia Magna of Industrial Organic Chemicals at Low Concentrations. Aquatic Toxicology, Vol. 12 p. 33-38.

Denton DL, Fox JF, Fulk FA. 2003. Enhancing toxicity performance by using a statistical criterion. Environ Toxicol Chem 22:2323-2328.

Felsot, A. 2005. A Critical Analysis of the Draft Report, "Basin Plan Amendments to the Water Quality Control Plan for the Sacramento River and San Joaquin River Basins for the Control of Diazinon and Chlorpyrifos Runoff into the Lower San Joaquin River"

Giddings JM, Soloman KR, Maund SJ. 2001. Probabilistic risk assessment of cotton pyrethroids: II. Aquatic mesocosm and field studies. Environ Toxicol chem 20 (3):660.

Host GE, Regal RR, Stephan CE. 1995. Analyses of acute and chronic data for aquatic life. Report United States Environmental Protection Agency, Washington, DC.

Hyder AH, Overmyer JP, Noblet R. 2004. Influence of developmental stage on susceptibilities and sensitivities of *Simulium vittatum* IS-7 and *Simulium vittatum* IIIL-1 (Diptera: Simulidae) to chlorpyrifos. Environ Toxicol Chem 23: 2856-2862.

Johnson WW, Finley MT. 1980. Handbook of acute toxicity of chemicals to fish and aquatic invertebrates. United States Fish and Wildlife Service Publication 137.

Lu, Zhimin and Gene Davis. 2009. Relative-Risk Evaluation for Pesticides Used in the Central Valley Pesticides Basin Plan Amendment Project Area. Central Valley Regional Water Resources Control Board. Sacramento, California. April 2008

Lydy MJ, Belden JB, Wheelock CE, Hammock BD, Denton DL. 2004. Challenges in regulating pesticide mixtures. *Ecology and Society* 9(6):1.

Mayer FL, Ellersieck MR. 1986. Manual of acute toxicity: interpretation and data base for 410 chemicals and 66 species of freshwater animals. United States Department of the Interior. Fish and Wildlife Service Resource Publication 160.

Mayes M, Weinberg J, Rick D, Martin MD.1993. Chlorpyrifos: A Life cycle Toxicity Test with the Fathead Minnow, *Pimephales promelas* Rafinesque: Lab Project Number; ES-DR-0043-4946-9: DECO-ES-2557B. Unpublished study prepared by The Environmental Toxicology & Chemistry Research Lab. 108p. MRID 428344-01

McCann, J. (1979) Dursban Daphnia magna 21 Day Life Cycle: Biological Report of Analysis: Static Test #2405. Unpublished study prepared by U.S. Environmental Protection Agency. 1 p. MRID 41073401

McClure, D., G. Davis, J. Karkoski, Lee, P. 2006. Amendments to the Water Quality Control Plan for the Sacramento River and San Joaquin River Basins for the Control of Diazinon and Chlorpyrifos Runoff into the Sacramento and San Joaquin Delta, Final Staff Report. California Regional Water Quality Control Board, Central Valley Region. Sacramento, CA.

RIVM. 2001. Guidance document on deriving environmental risk limits in The Netherlands. Report National Institute of Public Health and the Environment.

Siepmann S, Finlayson B. 2000. Water quality criteria for diazinon and chlorpyrifos. Report California Department of Fish and Game.

Sokal RR, Rohlf FJ. 1995. *Biometry, the Principles and Practice of Statistics in Biological Research*. New York: W. H. Freeman and Company, New York, NY.

State Water Resources Control Board (SWRCB). 2004. Water Quality Control Policy For Developing California's Clean Water Act Section 303(d) List.

State Water Resources Control Board (SWRCB). 2007. 2006 CWA Section 303(D) List Of Water Quality Limited Segments Requiring TMDLs. Available at: http://www.waterboards.ca.gov/water_issues/programs/tmdl/docs/303dlists2006/epa/state_06_303d_reqtmdls.pdf

State Water Resources Control Board (SWRCB). 2008. Draft Staff Report - Water Quality Control Plan for Enclosed Bays and Estuaries – Part 1 Sediment. Sacramento CA. 18 July 2008.

Suter GWI, Barnthouse LW. 1993. Assessment concepts. In: Suter GWI, editor. *Ecological Risk Assessment*. Boca Raton, FL: Lewis Publishers. p 21-47.

TenBrook PL, Tjeerdema RS. 2006. Methodology for derivation of pesticide water quality criteria for the protection of aquatic life in the Sacramento and San Joaquin River Basins. Phase I: Review of existing methodologies. Final Report. Central Valley Regional Water Quality Control Board, Rancho Cordova, CA.

U.S. EPA. 1980. Ambient Water Quality Criteria for Hexachlorocyclohexane, EPA 440/5-80-054. Report United States Environmental Protection Agency, Washington D. C.

U.S. EPA. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses, PB-85-227049. United States Environmental Protection Agency, National Technical Information Service, Springfield, VA.

U.S. EPA. 1986a. Ambient water quality criteria for chlorpyrifos, EPA 440/5-86-005. Report United States Environmental Protection Agency, Washington, D. C.

U.S. EPA. 1986b. Quality Criteria for Water 1986. EPA 440/5-86-001. United States Environmental Protection Agency, Washington, D. C.

U.S. EPA. 1996. Ecological Effects Test Guidelines. OPPTS 850.1020: Penaeid AcuteToxicity Test. EPA 712–C–96–130.

U.S. EPA. 1996. Ecological Effects Test Guidelines. OPPTS 850.1020: Gammarid AcuteToxicity Test. EPA 712–C–96–130.

U.S. EPA. 2002a. Interim Reregistration Eligibility Decision for Chlorpyrifos EPA 738-R-01-007. Report. United States Environmental Protection Agency, Washington, D. C.

U.S. EPA. 2002b.Understanding and accounting for method variability in whole effluent toxicity application under the National Pollutant Discharge Elimination System Program. EDs: Denton DL, Fox J, Fulk FA, Greenwald K, Narvaez M, Norberg-King TJ, Phillips L. EPA/833R-00-003. Office of Water. Washington DC.

U.S. EPA. 2002c. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms, 4th edition, EPA-821-R-02-013. Report United States Environmental Protection Agency, Washington, D. C.

U.S. EPA. 2007a. Aquatic Life Ambiient Freshwater quality Criteria- Copper: 2007 Revision. EPA-822-R-07-001. Report. United States Environmental Protection Agency, Washington, D. C.

U.S. EPA. 2007b. Aquatic Life Benchmark Table. U.S. Environmental Protection Agency, Office of Pesticide Programs (U.S. EPA). Washington, DC. Available at: http://www.epa.gov/oppefed1/ecorisk_ders/aquatic_life_benchmark.htm. Accessed on 15 August 2007.

Van Straalen NM, Van Leeuwen CJ. 2002. European history of species sensitivity distributions. In: Posthuma L, Suter GWI, Traas TP, editors. *Species Sensitivity Distributions in Ecotoxicology*. Boca Raton, FL: Lewis Publishers, CRC Press. p 19-34.

Viant, M.R., Werner, I., Rosenblum, E.S., Gantner, A.S., Tjeerdema, R.S., Johnson, M.L. 2004. Correlation between stress protein induction and reduced metabolic condition in juvenile steelhead trout (*Oncorhynchus mykiss*) chronically exposed to elevated temperature. *Fish Physiology and Biochemistry*, 29:159-171.

Werner I, Linares-Casenave J, Van Eenennaam JP, Doroshov SI. 2007. The effect of temperature stress on heat-shock protein expression and development in larval green sturgeon (*Acipenser medirostris*). *Environmental Biology of Fishes*. 71:191-200

Wheeler JR, Grist EPM, Leung KMY, Morritt D, Crane M. 2002. Species sensititivity distributions: stat and model choices. *Marine Pollut Bull* 45:192-202.